COMPARISON BETWEEN HIGH – RESOLUTION AERIAL IMAGERY AND LIDAR DATA CLASSIFICATION OF CANOPY AND GRASS IN THE NESCO NEIGHBORHOOD, INDIANAPOLIS, INDIANA

Nan Ye

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Geography,
Indiana University

January 2014

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

____
Daniel P. Johnson Ph.D., Chair

____
Frederick L. Bein Ph.D.

____
Vijay O. Lulla Ph.D.

ACKNOWLEDGEMENTS

I would like to thank my committee members: Dr. Dan Johnson, Dr. Rick Bein and Dr. Vijay Lulla for their guidance, help and incredible patience through this thesis research project. I am very thankful for their willingness to share their time and knowledge with me.

I am also very grateful to many other members of the Department of Geography. I want to thank Dr. Jeff Wilson, Dr. Rudy Banerjee, Jim Stout, Steve King, Austin Stanforth, Amanda Kabat, Tony Anthony, Joyce Haibe and Kathy Kozenski for their constant support and help with my thesis.

Finally, I want to thank my family who has supported and helped me throughout my education and life.

Nan Ye

COMPARISON BETWEEN HIGH – RESOLUTION AERIAL IMAGERY AND LIDAR DATA WHEN USED TO CLASSIFY CANOPY AND GRASS IN THE NESCO NEIGHBORHOOD, INDIANAPOLIS, INDIANA

Urban forestry is a very important element of urban structures that can improve the environment and life quality within the urban areas. Having an accurate classification of urban forests and grass areas would help improve focused urban tree planting and urban heat wave mitigation efforts. This research project will compare the use of high – resolution aerial imagery and LiDAR data when used to classify canopy and grass areas. The high – resolution image, with 1 – meter resolution, was captured by The National Agriculture Imagery Program (NAIP) on 6/6/2012. Its coordinate system is the North American Datum of 1983 (NAD83). The LiDAR data, with 1.0 – meter average post spacing, was captured by Indiana Statewide Imagery and LiDAR Program from 03/13/2011 to 04/30/2012.The study area is called the Near East Side Community Organization (NESCO) neighborhood. It is located on the east side of downtown Indianapolis, Indiana. Its boundaries are: 65 interstate, East Massachusetts Avenue, East 21st Street, North Emerson Avenue, and the rail road tracks on the south of the East Washington Street. This research will also perform the accuracy assessment based on the results of classifications using high – resolution aerial imagery and LiDAR data in order to determine and explain which method is more accurate to classify urban canopy and grass areas.

Daniel P. Johnson Ph.D., Chair

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Urban forests are an important element and component within urban areas. They are highly correlated with quality of environment, which provides many environmental benefits including carbon sequestration, storm water interception, soil erosion control, habitat for wildlife, and energy conservation. Coder (1996) stated that trees can partially control pollutants including carbon dioxide, ozone, and small particulates less than 10 microns in size. Dwyer, McPherson, Schroeder, and Rowntree's research showed that the pine trees in Los Angeles are projected to remove from the atmosphere (under 400 meters) about 8% of the ozone and decrease the concentration around the leaves by 49% (Dwyer, McPherson, Schroeder, & Rowntree, 1992). Rowntree and Nowak stated that urban tree canopies could provide a great benefit of reducing carbon dioxide. If the United States could increase average urban tree cover by 5% (from 28% to 33%) over the next 50 years, the urban tree canopy could store an additional 150 million tons of carbon (Rowntree & Nowak, 1991).

As cities and urban areas grow rapidly, the growth makes significant impact on urban ecosystem conditions (like urban heat island, air, and water qualities). In order to develop strategies for improving the environment of cities and urban areas, accurate and updated information of the status and trends of urban ecosystems is needed (Yang, Xian, Klaver, & Deal, 2003).

In the last decade, high – resolution satellite sensors became available, which allowed a higher resolution level monitoring. For example, GeoEye – 1 now provides 0.41 – meter (16 – inch) panchromatic and 1.65 m (5.4 feet) multispectral imagery. With the development of satellite and remote sensing technology, there are more and more advanced strategies and means to monitor changes of urban landscapes. Usually, high – resolution commercial satellite imageries are very expensive to access. Aerial imageries also have a decent resolution quality (about 1 meter). But both high – resolution satellite imageries and aerial imageries have some limitations for classifying urban areas. Their first four bands, including band 1(blue), band 2(green), band 3(red), and band 4(near – infrared), could have 1 meter and 1.65 meter resolution. But in comparing them to the great difference within urban areas, such as buildings, sidewalks, pavements, tree canopies, grass space, and many other land features, even a 1 meter size pixel could contain a lot of information, which would reduce accuracy of the image classification. Shading areas, caused by tall buildings, trees, topography, and many other items, will also reduce accuracy of the image classification. Especially, when it's focused on a small study area, the shading areas could take a high percentage of the whole study area.

LiDAR stands for Light Detection and Ranging that uses laser pulses to gauge how far an object on the ground is from the sensor (Harrap & Lato, 2006). LiDAR is another technology for rapidly measuring positions of land features, especially in urban areas (Harrap & Lato, 2006). Comparing LiDAR to remote sensing imagery and aerial imagery classification, there are many advantages of using LiDAR for urban forestry applications. For example, Palmer and Shan stated that LiDAR can allow to rapidly acquire a large – scale DTM (digital terrain model); LiDAR is daylight independent, weather independent, and highly precise (Palmer & Shan, 2002). It can also generate 3D topographic surface information more rapidly than many other technical methods (Shan & Sampath, 2005). LiDAR provides the number of signal returns that can retrieve the vegetation height information for the urban canopy classification (Hodgson, Jensen, Tullis, Riordan, & Archer, 2003).

The purpose of this study is to evaluate and compare the methods of urban tree canopies and grass space classification by using aerial imagery and LiDAR data. Technically, the result of classification using LiDAR data is assumed to be better than using aerial imagery. By using LiDAR data, the classification will be available for obtaining the elevation data of all the land features (such as buildings, tree canopies) that would help with identifying tree canopies more accurately. This study will focus on the NESCO neighborhood of Indianapolis, IN.

BACKGROUND

As people focus more on public health, the research of urban forestry has become popular. Coder's (1996) research showed that there are many environmental benefits from urban forests such as increasing shading areas, wind control, pollution reduction, and carbon dioxide reduction. Coder (1996) also stated that the urban forests can provide economic and social benefits such as increasing property values and providing visual screens to block vision lines that show human density problems or residential interfaces.

With the higher GIS technology being developed to benefit the human society, the urban forest, and vegetation management starts to rely on the tool of GIS more and more. Even though the traditional satellite image classification and aerial photo classification has made a huge contribution to the natural resource classification, there are not many accurately well – developed methods for urban plantable areas (grass area) and tree canopy classification.

This research will use GIS tools (ERDAS IMAGINE and ArcGIS) to do the classification analysis in NESCO Indianapolis. The research project for KIB (Keep Indianapolis Beautiful INC.) was to help KIB find out the most efficient GIS model and decide which areas are the best and most appropriate to plant trees. KIB project classified an aerial image of NESCO area into 5 classes that are plantable area (grass), tree canopy, impervious surfaces, shadows, and unknown areas. There are several geoprocessing tools used in ArcMap to figure out the plantable area and tree canopy area within NESCO. Finally, the data was published into the ArcGIS Server and a web application service was built up for KIB and the public to see and use the maps to help them make decisions on where to plant trees. The results of plantable areas and tree canopy classifications are appropriate, but there are still some problems and limitations that reduce accuracy of classification. For example, shading caused by tree canopies and tall buildings will reduce classification accuracy; the resolution problem: when it's focused on a small area to do a classification, high resolution is needed. The aerial imagery we used is 1 – meter resolution. When focused on an urban area where much variance is included, it will still be limited to displaying the details. LiDAR data, with its embedded elevation information, could prove to be a more accurate basis for the classification of grass and canopy.

Study Area (NESCO) Background

The study area in the research is called NESCO (Near East Side Community Organization). NESCO was first organized in the spring of 1971. The boundaries are: 65 interstate, East Massachusetts Avenue, East 21st Street, North Emerson Avenue, and the rail road tracks in the south from East Washington Street. The total area of NESCO is about 5.75 square miles.

Figure 1: Logo of NESCO.

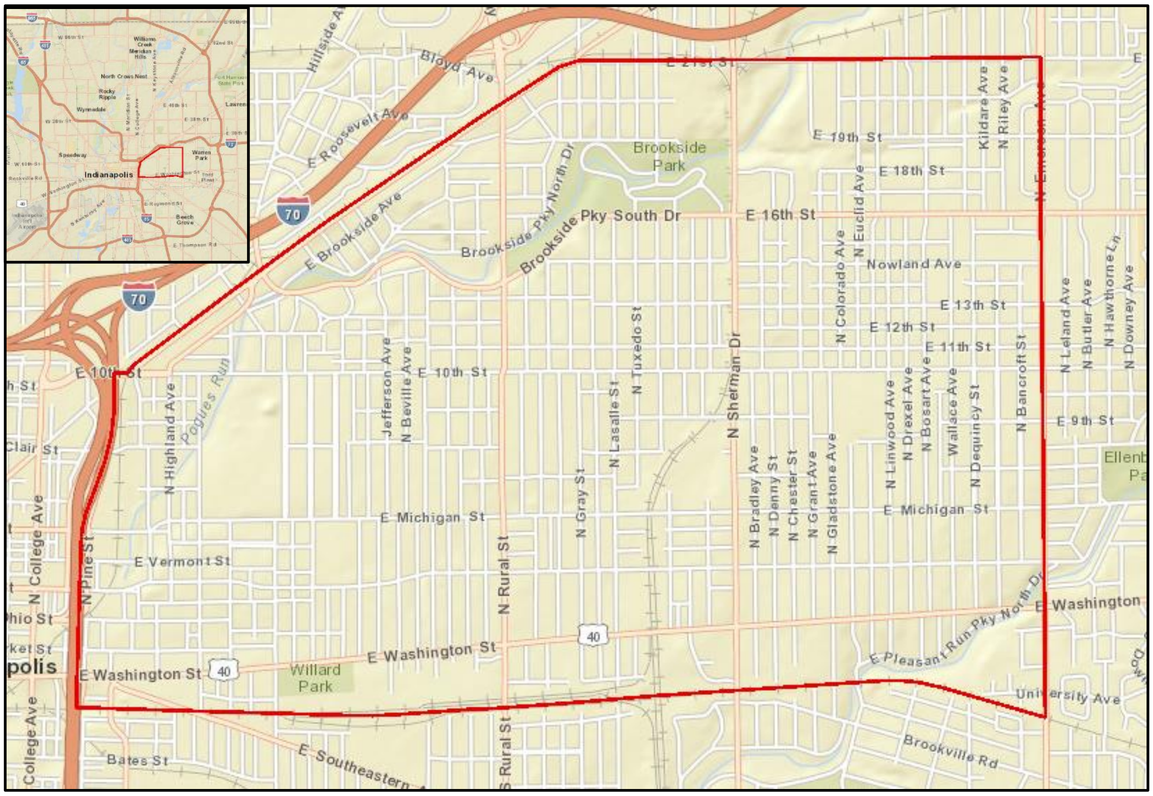


Figure 2: NESCO Boundary.

Remote Sensing and Image Classification

Remote sensing is a way of collecting information about objects or phenomenon without making any physical contact with them. The modern technology of remote sensing includes both satellite and aerial based systems. It helps and allows us to collect lots of physical data quick and easily. To combine using remote sensing and GIS (Geographical Information Science), it helps us analyze the data spatially and offer possibilities of producing various options that could optimize the entire planning process (Verma, Kumari, & Tiwary, 2009).

The development of the satellite has promoted the high – resolution data in remote sensing. For example, Landsat 8 that was launched February 11th, 2013 can now provide 15 – meter / 30 – meter / 100 – meter (panchromatic/multispectral/thermal) imageries. GeoEye – 1 satellite was launched September 6th, 2008 that provides 0.41 – meter panchromatic and 1.65 – meter multispectral imageries. This study will use aerial images accessed from the Indiana Spatial Portal (http://gis.iu.edu/) and the resolution of the images used is 1 – meter with 4 bands (Blue, Green, Red, and Near Infrared).

Unsupervised Classification

The imagery classification is divided into two: supervised and unsupervised. The main difference between these two methods is how the spectral signatures are created (Kokalj & Kri, 2007). Unsupervised classification is also called clustering whose most widely used algorithms include K Means and isodata standing for iterative self – organizing data analysis technique (Ari & Aksoy, 2010). Clustering is a method of collecting likeness of data, making a group based on the similarity and dissimilarity and a classified group is a cluster (Arai & Bu, 2007).

Supervised Classification

Besides unsupervised classification, there is another essential tool used for extracting quantitative information from remote sensing imagery data that is supervised classification. Liu (2008) stated that supervised classification uses sufficient enough groups of known pixels to generate classes of interest. Each group of known pixels is called a "Training Area." Each generated class has the same or similar value with a relevant group of known pixels. Despite the advantages of supervised classification, there are few limitations to it. For instance, information classes may not match spectral classes, which mean there would be pixels unclassified. Training areas may not include unique spectral class or classes, which would also leave pixels unclassified.

The thesis will talk about one most commonly used supervised classification that is maximum likelihood classification (MLC). Maximum likelihood classification is based on Bayes' theorem. It uses a discriminant function to assign pixels to the class with the highest likelihood and the class mean vector and covariance matrix are the key inputs to the function (Ahmad & Quegan, 2012). These inputs can be calculated from the training areas.

LiDAR Classification

LiDAR stands for Light Detection and Ranging system. It contains three main technologies: Lasers, Global Positioning System (GPS), and Inertial Navigation System (INS), which can help to be capable of acquiring data to produce highly accurate Digital Elevation Models (DEMs).

There are many essential advantages of using LiDAR data including the ability to go through a tree canopy, detect its vertical information, and quantify its structure (Gatziolis & Andersen, 2008). Another advantage of using LiDAR is that LiDAR data can be collected under many different environmental conditions like a low sun angle (in other words, no shadow influence), cloudy weather, and even darkness while the satellite sensors have those limitations to collect spatial data (Veneziano, Hallmark, & Souleyrette, 2002).

By using multi – return LiDAR, it will be able to collect dense point data reflected by the first surface (canopy) and go through the vegetation cover with many points hitting the ground (Renslow, Greenfield, & Guay, 2000). With this multi – return technology, we are able to get the information about the vegetation vertical profile very easily and accurately.

Collection of elevation data using LiDAR has several advantages over most other techniques. Most among them are higher resolutions, higher accuracies, and ground detection under forests.

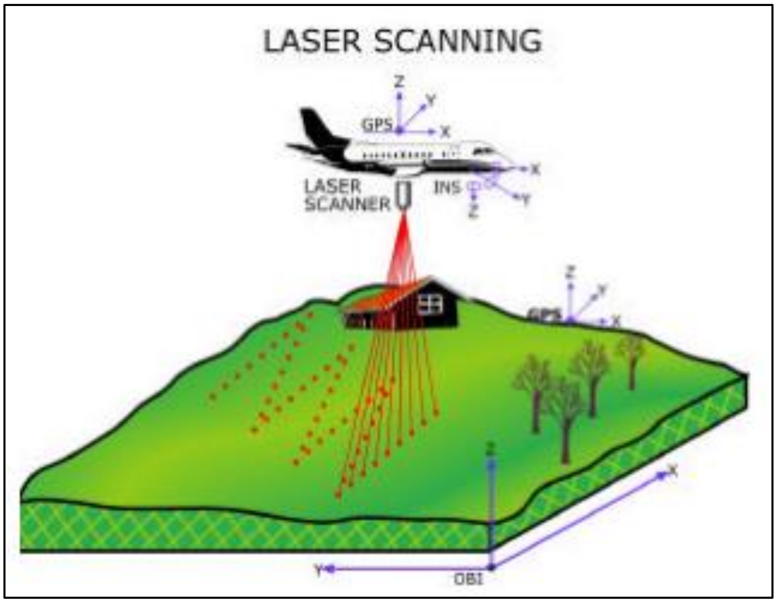


Figure 3: Operational Characteristics of LiDAR Data Collection.

The picture above is retrieved from http://www.fs.fed.us/pnw/olympia/silv/lidar/, 2/10/2013. The red dots represent LiDAR points hitting the ground at a specified post – spacing, in a wave – like scanning pattern.

Liu (2008) stated that the LiDAR system is able to detect multiple return signals of transmitted pulse, and most LiDAR systems could detect first and last returns while some others are able to detect up to 6 returns for a single pulse.

Liu's (2008) study also showed that the multiple returns usually happen when a laser pulse hits a target that does not totally block the path of the pulse and the remaining portion of pulse could continue to go to a lower position of the target. The multiple returns often happen in an area with canopies, because there are many gaps between branches and leaves. Therefore, by using the multiple returns we can easily identify the canopy area.

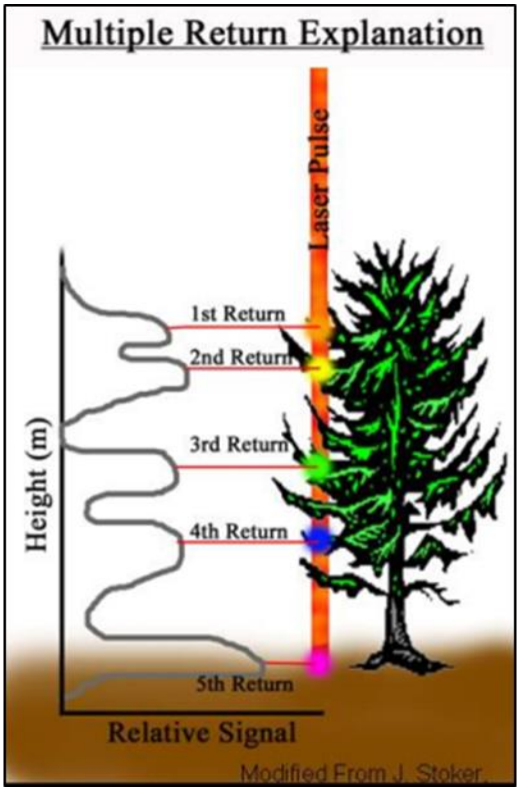


Figure 4: Multiple Returns from Single Pulse (National Oceanic and Atmospheric Administration (NOAA), 2012).

Here is an example in Figure 4 (retrieved from http://www.csc.noaa.gov/digitalcoast/_/pdf/lidar101.pdf, 2/10/2013). There could be 5 returns from single laser pulse. The first return is reflected from top of the canopy. And the returns from the second return through the fourth return are reflected under the canopy while the fifth return is reflected from the ground. Therefore, the returns from first return to fourth return represent the tree and the last return represents ground.

The most common LiDAR point cloud file format is the LAS format. The LAS file format is a public file format for the interchange of 3 – dimensional point cloud data. Isenburg stated that it is now the industry standard for storage and distribution of airborne LiDAR data (Isenburg, 2013). The latest version of LAS format is 1.4. In this research, the LAS format data I am going to use will be version 1.0.

There is an American Society for Photogrammetry and Remote Sensing (ASPRS) classification scheme (retrieved from http://www.csc.noaa.gov/digitalcoast/_/pdf/lidar101.pdf, 2/10/2013) that is used by most LiDAR producers (Figure 5). The National Oceanic and Atmospheric Administration (NOAA) (2012) stated that the primary values are 1, 2, and 9.

Classification Value and Meaning
0 Created, never classified
1 Unclassified
2 Ground
3 Low Vegetation
4 Medium Vegetation
5 High Vegetation
6 Building
7 Low Point (noise)
8 Model Key-point (mass point)
9 Water
10 Reserved for ASPRS Definition
11 Reserved for ASPRS Definition
12 Overlap Points
13-31 Reserved for ASPRS Definition

Figure 5: ASPRS Classification Values.

Intensity is another important component of LiDAR data that is the energy reflection from objects on the earth surface recorded by the sensor (Singh, Vogler, Meng, & Meentemeyer, 2010). There are four main factors affecting intensity measurements that are surface reflectance, atmospheric transmission, local incidence angle, and the distance between the sensor and earth objects (Singh, Vogler, Meng, & Meentemeyer, 2010). Potentially, Intensity could help with grass classification effectively.

LiDAR technology provides horizontal and vertical information at high spatial resolutions and vertical accuracies, and therefore LiDAR technology is an efficient and accurate way to classify urban forestry. LiDAR technology can directly retrieve the forest attributes such as canopy height (Lim, Treitz, Wulder, St-Onge, & Flood, 2003). Typically, satellite imageries and aerial photos are only capable to provide detailed information on the horizontal distribution without the vertical distribution of forests.

However, LiDAR data provides both horizontal distribution and vertical distribution that can help accurately classify the vegetation in forests such as canopies (Lim, Treitz, Wulder, St-Onge, & Flood, 2003).

METHODOLOGY

In this section, the research project will explain three methods of classification processes. These are unsupervised classification (100 – class classification), supervised classification, and LiDAR classification.

Imagery

This aerial imagery below was downloaded from Indiana Spatial Data Portal (gis.iu.edu, 2/20/2013). The imagery is provided by The National Agriculture Imagery Program (NAIP) and was captured on 6/6/2012 with 1 – meter resolution. The NAIP acquires 4 – band digital aerial imagery from airborne and/or space based platforms during the agricultural growing seasons in the U.S. The NAIP imagery is formatted to the UTM coordinate system using the North American Datum of 1983 (NAD83). Figure 6 is a clip of the aerial imagery containing the IUPUI campus, the downtown of Indianapolis, IN, and NESCO.

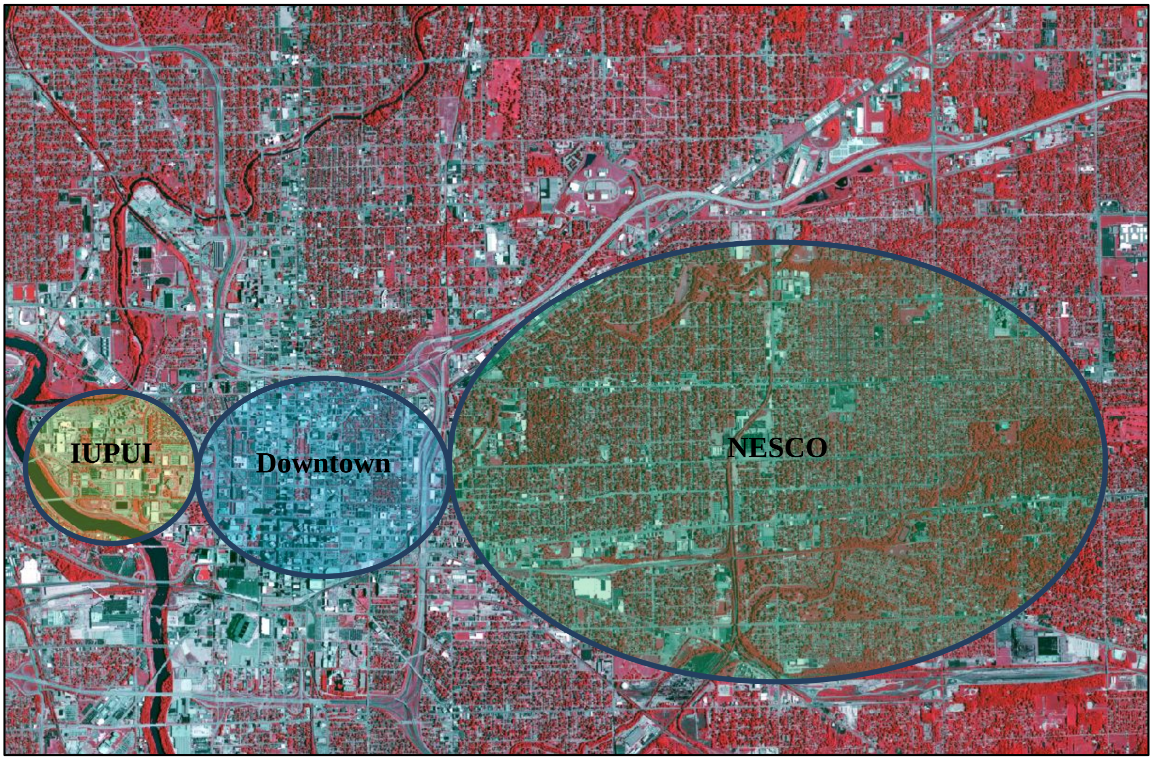


Figure 6: NAIP Image (July 2012) in NESCO.

LiDAR Data

The LAS data for this study area is downloaded from http://www.indianamap.org/ 6/5/2013. It was taken from 03/13/2011 to 04/30/2012 with a resolution of 1.0 – meter average post spacing. Indiana's Statewide LiDAR data is produced for all 92 Indiana Counties covering more than 36,420 square miles (Indiana Statewide Imagery and LiDAR Program, 2013).

The LAS file cannot be directly shown in ArcMap. Usually it needs to be converted into an LAS dataset to be shown in ArcMap. This figure below is an area view of the NESCO LAS dataset. Basically, they are elevation points.

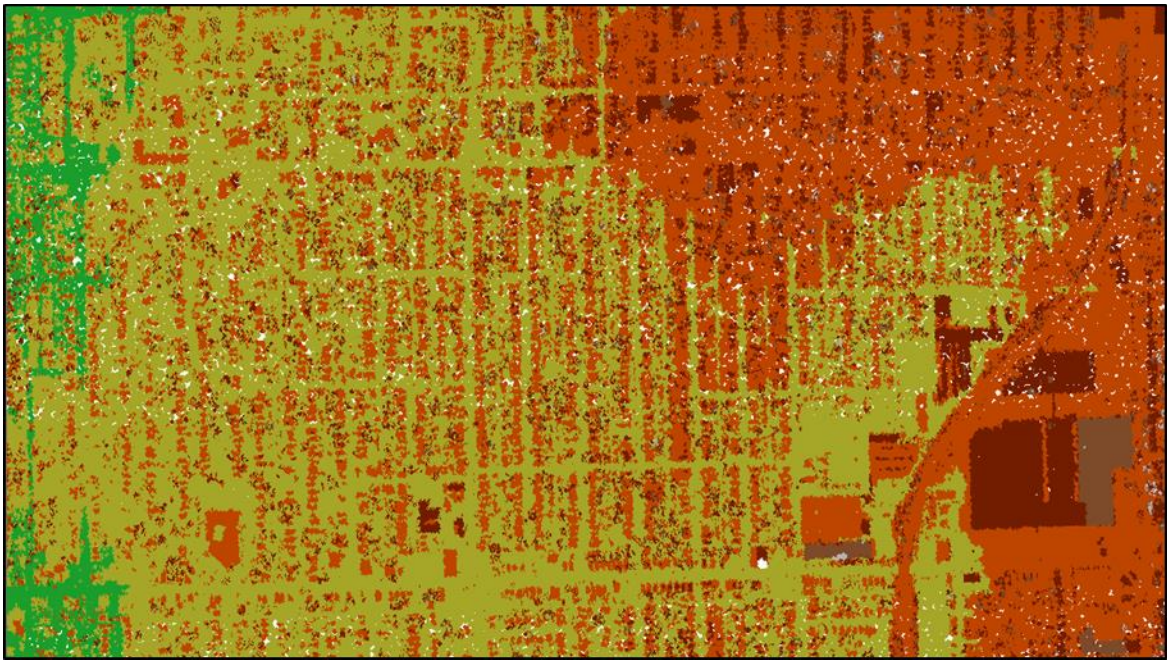


Figure 7: NESCO LiDAR Dataset Shown in Elevation Mode (December 2009).

Unsupervised Classification (100 Classes)

The K Means method is a non – hierarchical clustering method proposed by MacQueen, Anderberg, Forgy, and others in the 1960s (Arai & Bu, 2007).

The Figure 8 (retrieved from http://www.cs.uvm.edu/~xwu/kdd/Slides/Kmeans – ICDM06.pdf, 6/17/2013) below shows the demonstration of K Means.

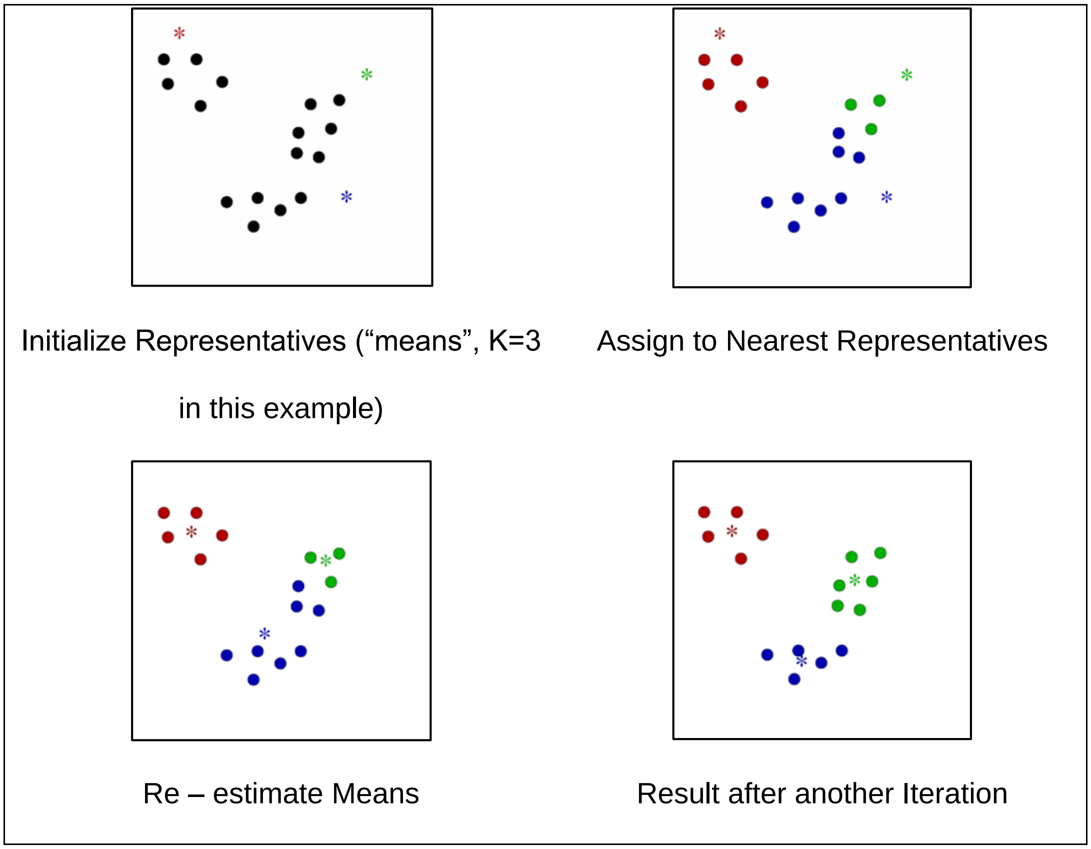
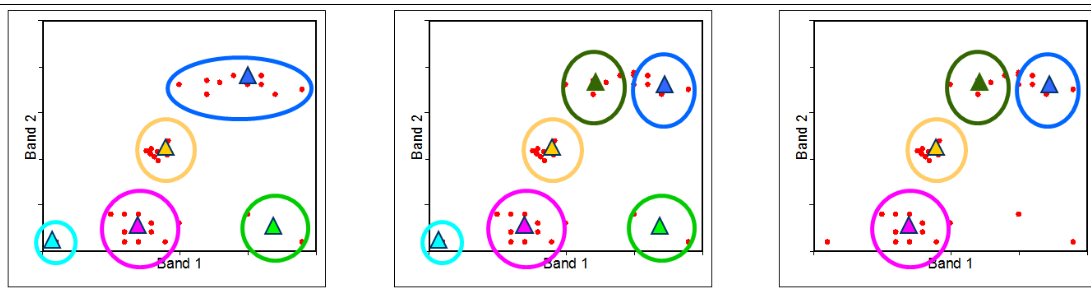


Figure 8: Demonstration of K Means.

K Means also has some limitations and flaws. Because K Means only focuses on all the distance of individuals and their centers, it can guarantee local optimal nature, but not global optimal nature. Another flaw is that the result of clustering will change when the K value is different or the initial cluster centers setup is changed (Arai & Bu, 2007).

ISODATA stands for iterative self – organizing data analysis technique that is another very popular method of unsupervised image classification that only needs three input parameters: the number of classes (clusters), the maximum number of iteration, and the convergence threshold, that basically means the maximum percentage of pixels whose class values are allowed to be unchanged between iterations (Al-Ahmadi & Hames, 2009).



Data is clustered but blue cluster is very stretched in band 1	Cyan and green clusters only have 2 or less pixels. So they will be removed.	Either assign outliers to nearest cluster, or mark as unclassified.

Figure 9: Demonstration of ISODATA.

ISODATA classification method is more advanced than K – means, since it has division performance of clusters. The global optimal nature cannot still be guaranteed by ISODATA (Arai and Bu, 2007).

This research will use ISODATA for unsupervised classification as an example. Figure 10 shows the setting box for ISODATA in ERDAS IMAGINE software. There are 100 classes set in this process. The maximum iterations are set with 24 and the convergence threshold is set with 0.950. Imagine Tour Guides V8.4 (1999) stated that 24 for maximum iterations is good in practical applications and the maximum number of iterations prevents the re – clustering process from running too long, or from potentially getting stuck. Imagine Tour Guides V8.4 (1999) also stated that the convergence threshold is the maximum percentage of pixels whose cluster assignments will not change between iterations. 0.950 for the convergence threshold means that as soon as 95% or more of the pixels stay in the same cluster between one iteration and the next, the ISODATA process will stop processing. Figure 11, 12, and 13 show the results of 100 – class classification.

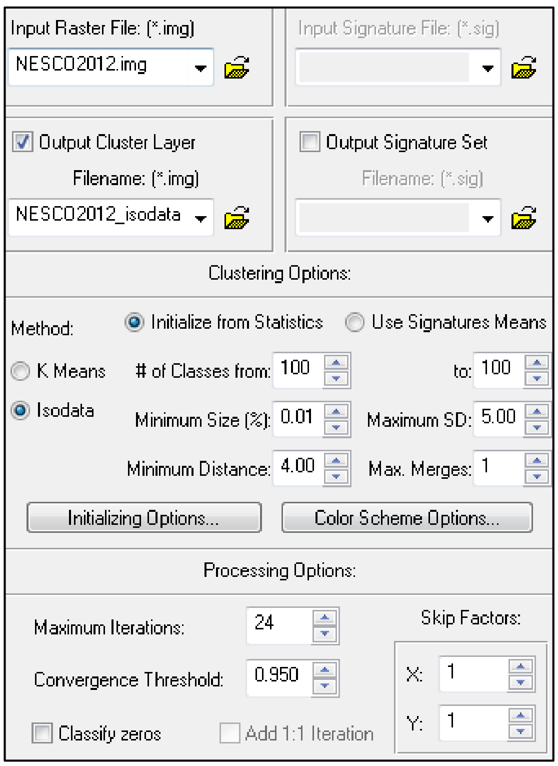


Figure 10: Settings for ISODATA in ERDAS IMAGINE.

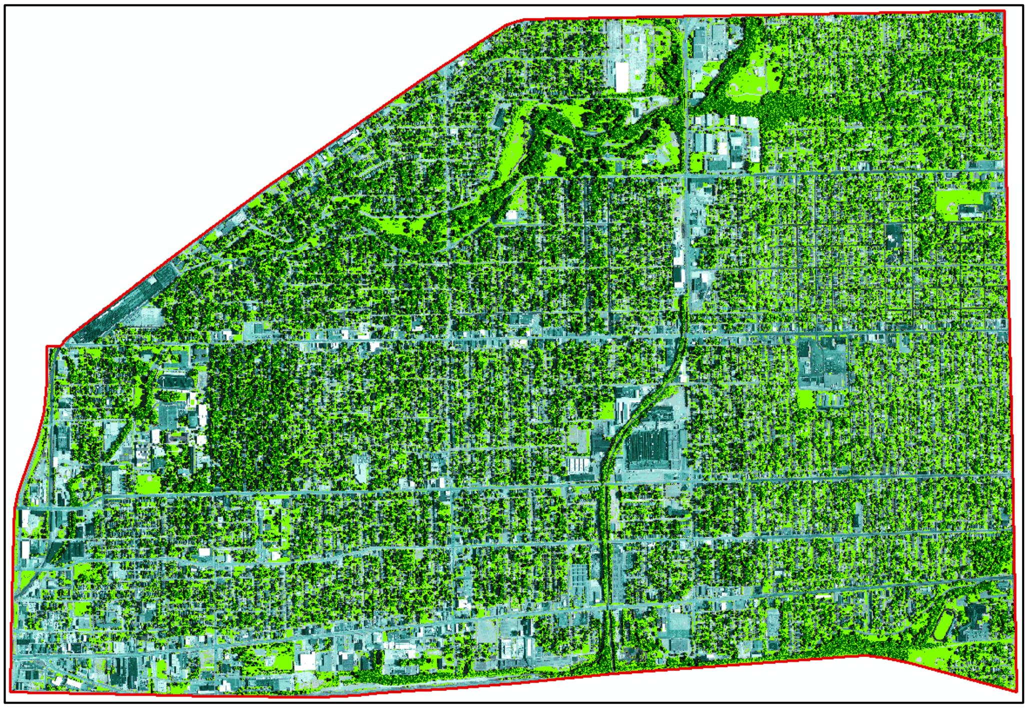


Figure 11: 100 – Class Classification.

Figure 12: 100 – Class Classification (Dark Green: Canopy, Light Green: Grass, Black: Unclassified).

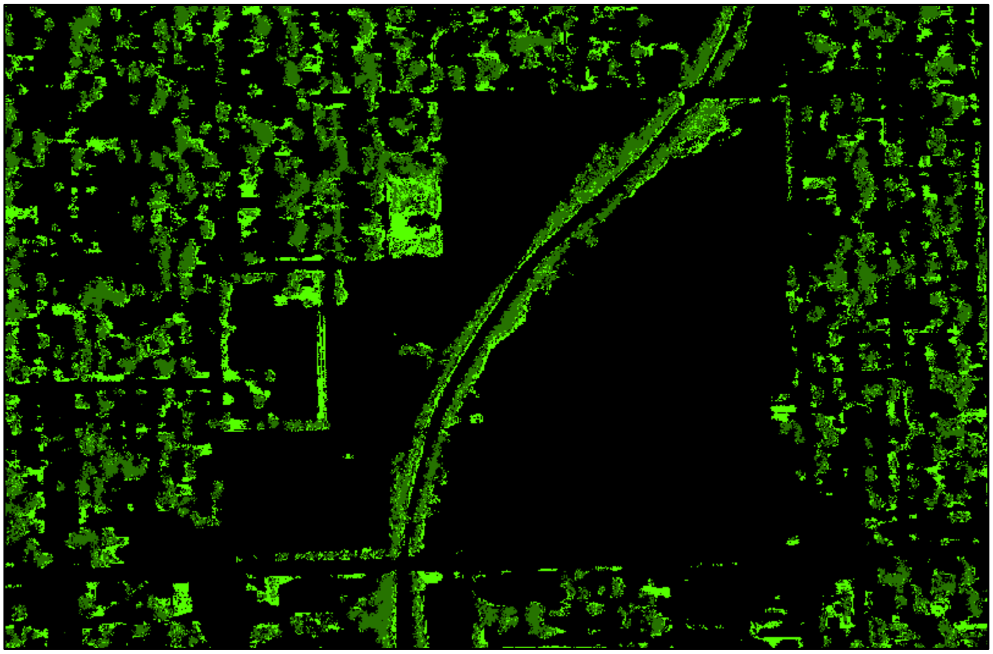
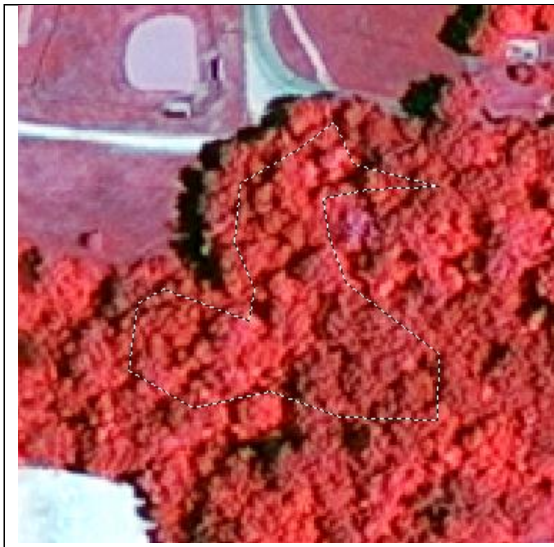
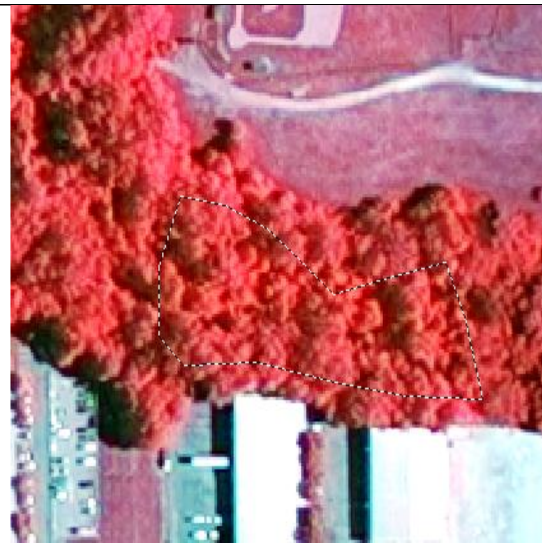
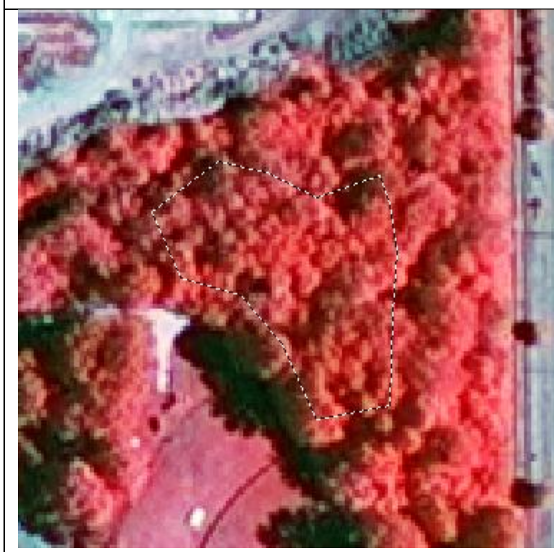
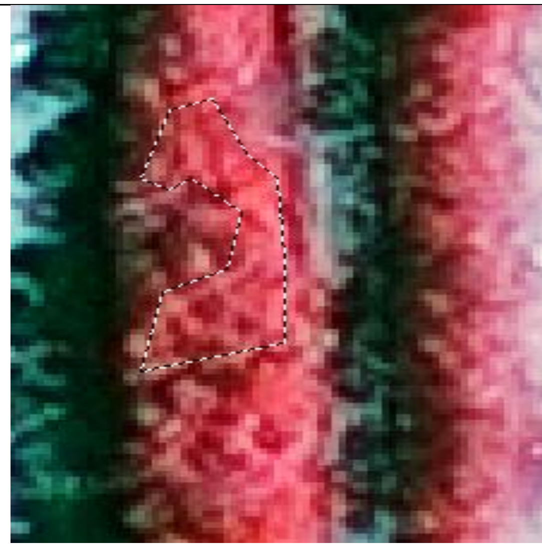


Figure 13: 100 – Class Classification (Dark Green: Canopy, Light Green: Grass, Black: Unclassified) in Detail.

Supervised Classification

In this section on supervised classification, the process will randomly pick up 10 training areas for each canopy classes and grass classes. The decision rules for supervised classification will be that non – parametric rule: parallelepiped, overlap rule: parametric rule, unclassified rule: unclassified, and parametric rule: maximum likelihood. Here are the 10 training areas for canopy classes shown below.

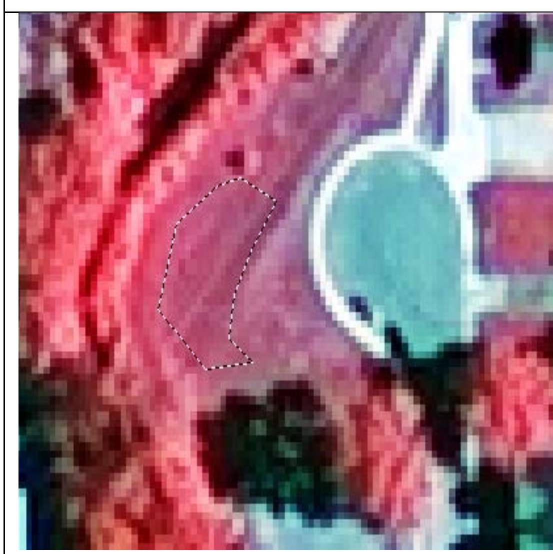
	
Training area 1 for canopy	Training area 2 for canopy
	
Training area 3 of canopy	Training area 4 of canopy

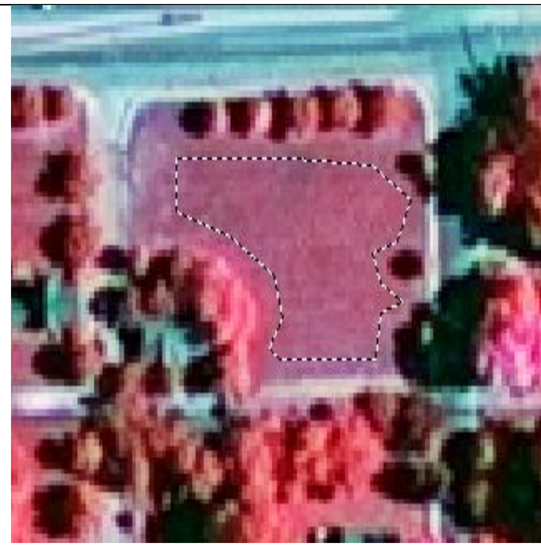
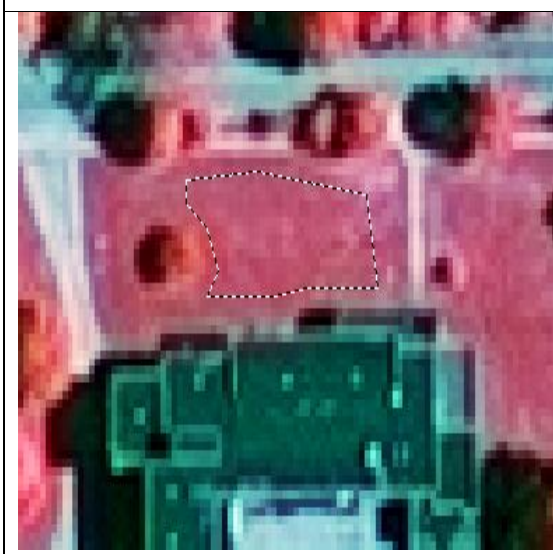
	
Training area 5 of canopy	Training area 6 of canopy
	
Training area 7 of canopy	Training area 8 of canopy

	
Training area 9 of canopy	Training area 10 of canopy

Table 1: 10 Training Areas for Canopy Classes.

Here are the 10 training areas for grass classes shown below.

	
Training area 11 of grass	Training area 12 of grass
	
Training area 13 of grass	Training area 14 of grass

	
Training area 15 of grass	Training area 16 of grass
	
Training area 17 of grass	Training area 18 of grass

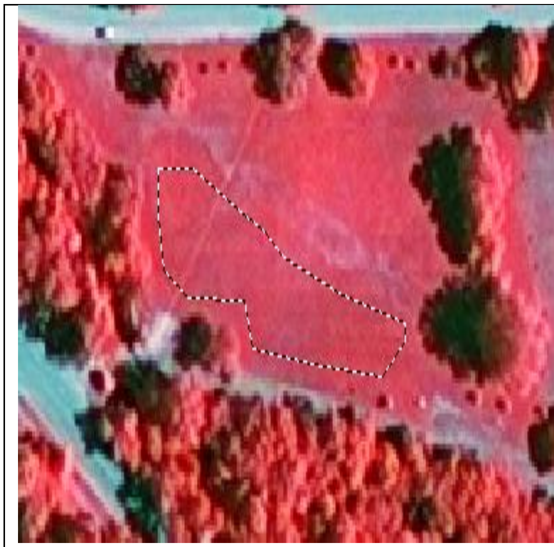
	
Training area 19 of grass	Training area 20 of grass

Table 2: 10 Training Areas for Grass Classes.

Signature Name	Color	Number of Pixels (Size 1 x 1 Meter^2)
Class 1		5866
Class 2		5482
Class 3		3275
Class 4		902
Class 5		1714
Class 6		610
Class 7		602
Class 8		242
Class 9		394
Class 10		363
Class 11		442
Class 12		235
Class 13		320
Class 14		155
Class 15		422
Class 16		1277
Class 17		456
Class 18		294
Class 19		2200
Class 20		2235

Table 3: Signatures of 20 Classes.

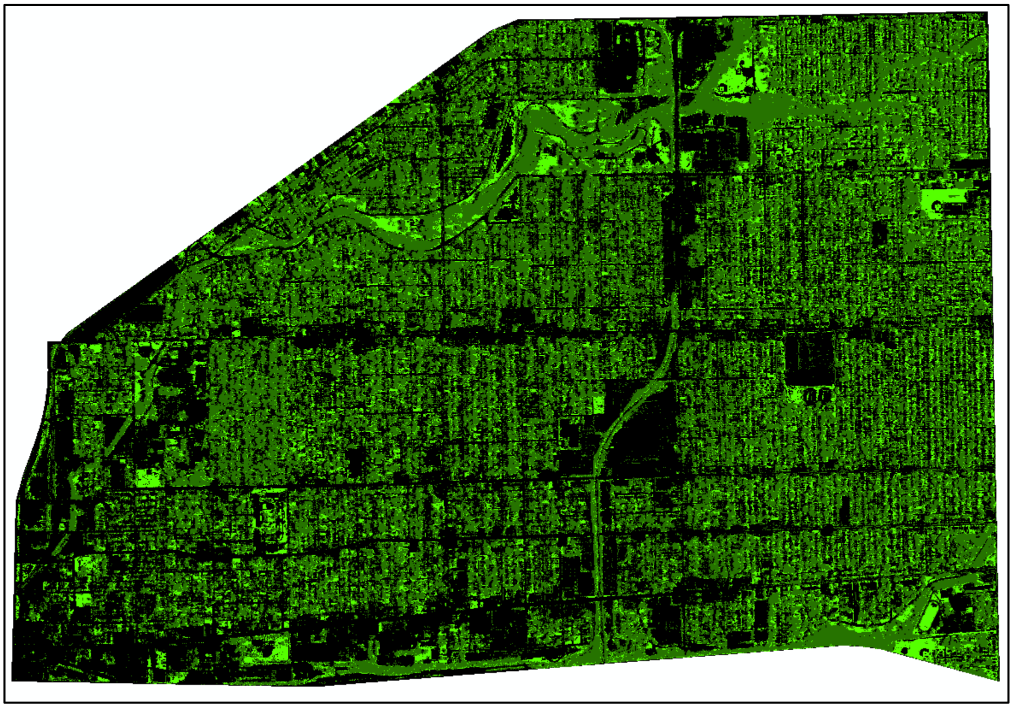


Figure 14: Supervised Classification (Light Green: Grass, Dark Green: Canopy, Black: Unclassified).

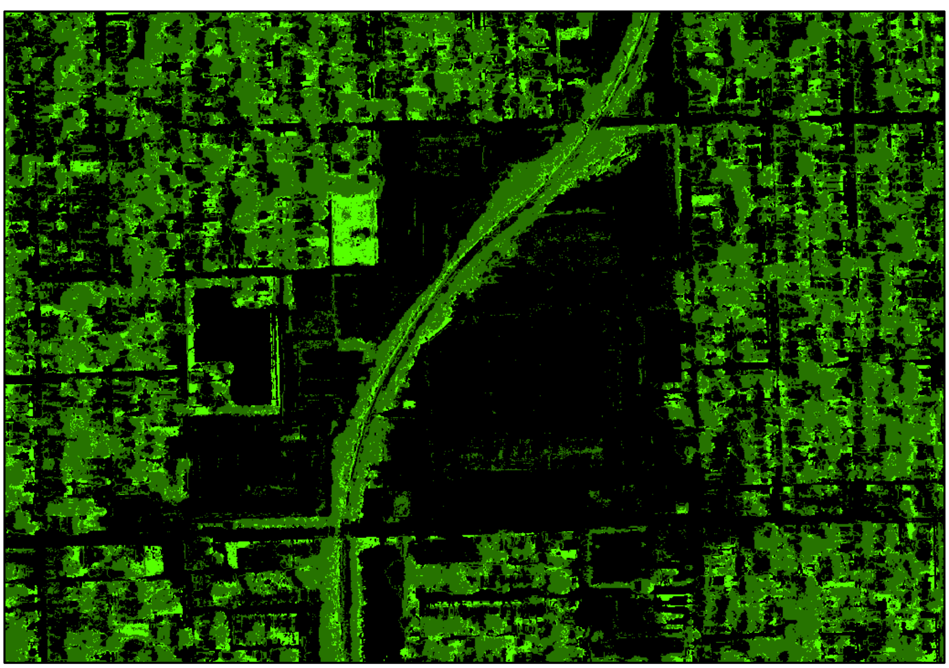


Figure 15: Supervised Classification (Light Green: Grass, Dark Green: Canopy, Black: Unclassified) in Detail.

Classification Using LiDAR data

The version of the LAS data for this research project is 1.2.

Classification Codes

Classification	Point Count	%	Z Min	Z Max	Min Int...	Max Int...	Synthe...
1 Unassigned	11,195,404	26.65	207.55	382.43	0	255	0
2 Ground	13,740,165	32.70	214.82	252.20	0	255	0
7 Noise	41,280	0.10	190.65	250.90	0	255	0
9 Water	9,354	0.02	233.81	236.88	0	240	0
10 Reserved	193,396	0.46	218.51	253.49	0	255	0
12 Overlap	16,833,299	40.07	203.64	382.65	0	255	0

Table 4: Classification Codes Table.

As the canopy and grass areas are not classified in this dataset, this research will manually classify the canopy and grass areas using LiDAR LAS dataset.

Here is the general classification process in this methodology section. Firstly, classify tree canopy areas (by using different signal returns). Secondly, classify all the land features (buildings and canopy) above the ground level (by using Digital Elevation Model). Thirdly, classify grass areas (by intensity values). Fourthly, remove some grass errors on pavement and buildings by using DEM and pavement shapefile. Lastly, add all the layers together and get the result layer.

LiDAR Classification Process

Step 1: Classify canopy areas

Using multiple return points allows the capture of three – dimensional structure of the “features above the ground surface,” such as the forest canopy and understory. In ArcMap, from the dataset property window, the layer was set only to show “first of many” and “last of many” returns which represent canopy generally (Figure 16). Generating a raster from the LiDAR dataset enables a layer of tree canopy (Figure 17, 18, 19, and 20).

Step 2: Classify all land features above ground (buildings and tree canopy).

The reason for not only getting the buildings layer is because in the LiDAR dataset some of the tree canopies and buildings are overlapped. The DEM (Digital Elevation Model) layer (Figure 21 and 22) and the DTM (Digital Terrain Model) layer (Figure 23, 24, and 25) can easily be converted from the LiDAR dataset.

Since the ground level is not flat, using the layer of DEM with the features above the ground level minus ground level layer can deliver all the features that are above the ground level and based on the same flat level (Figure 26, 27, and 28).

Step 3: Classify grass areas (by intensity values)

The intensity information stored in the LiDAR dataset can reveal a raster based on the intensity value (Figure 29 and 30). In Figure 30, the intensity values of canopies, most roofs, pavements, and many other impervious surfaces like drive ways are lower than grass areas. Figure 31 and 32 show the intensity raster with different display.

Therefore, choosing high intensity value pixels would help find the grass surfaces (including some other few rough surfaces) (Figure 33 and 34). In Figure 34, some impervious surfaces like buildings roofs and pavements are covered by the high intensity values too. So pavements shapefile provided by IMAGIS (Figure 35) and the raster layer of all features above bare ground (Figure 27) can remove the errors on roofs and pavements. Figure 36 and 37 show the high intensity value raster with all the features above bare ground and pavements removed, which is the grass areas.

Step 4: Add All the Layers Together and Get the Result Layer

Figure 38 and 39 show the result layer with canopy, grass classes.

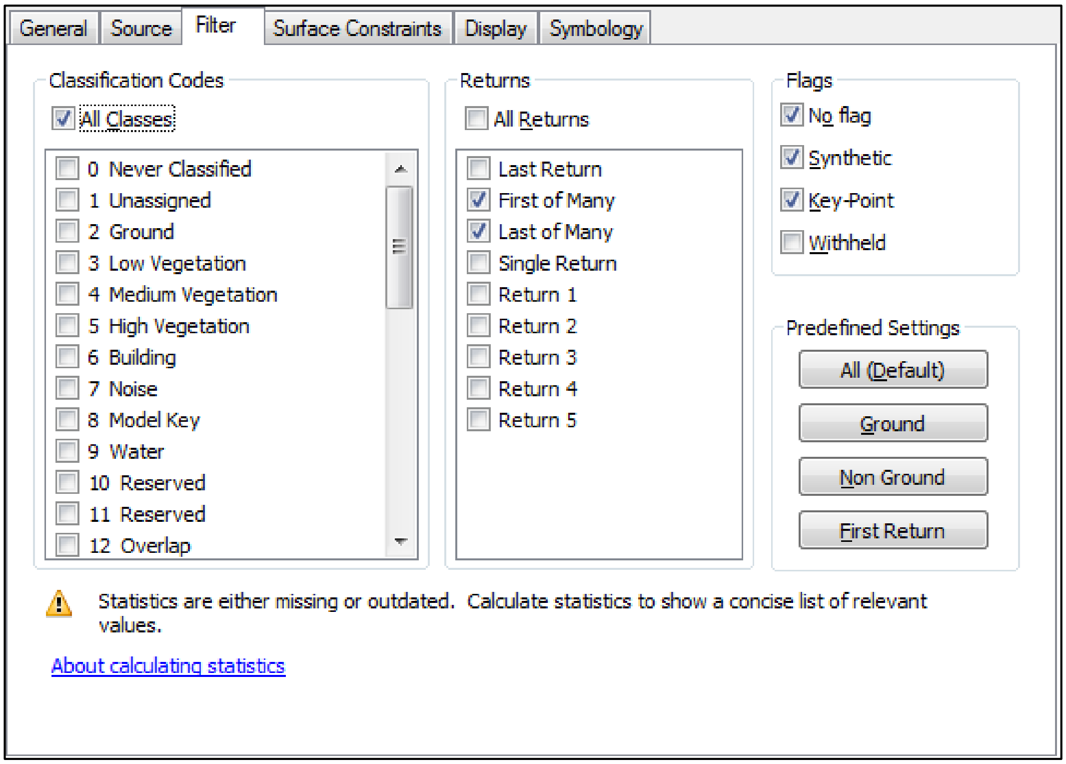


Figure 16: Table of Returns.

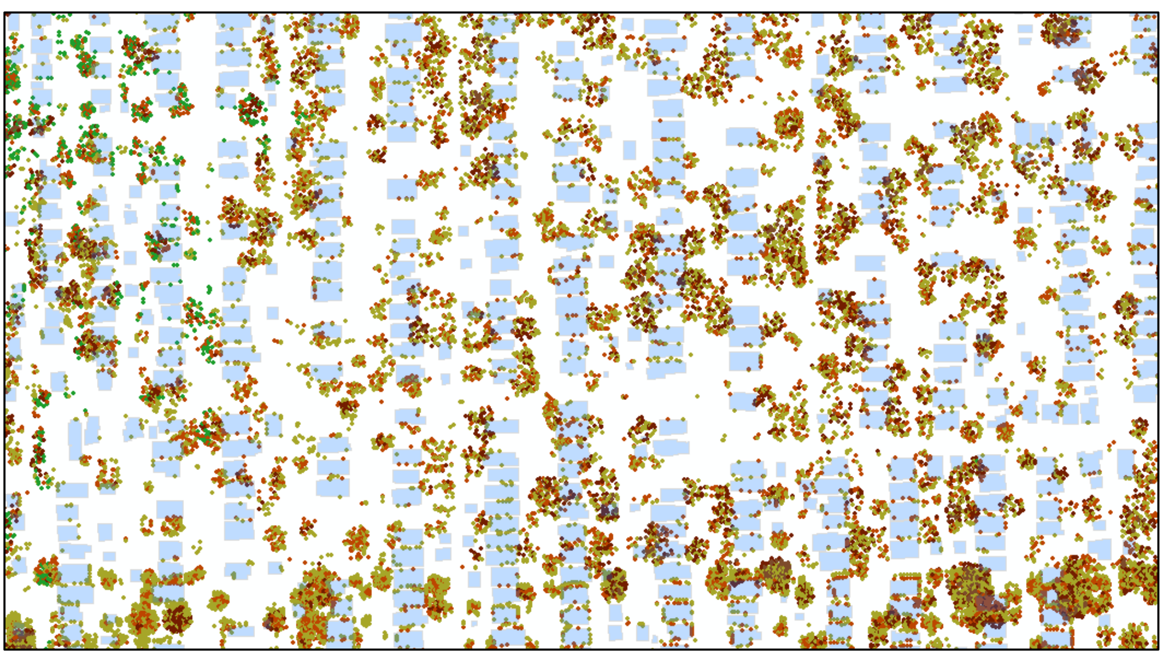


Figure 17: Canopy and Buildings Shapefile.

ArcTool – LAS Dataset to Raster.

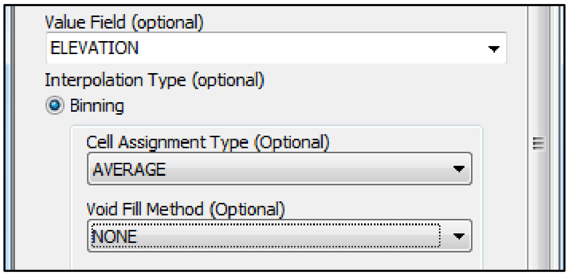


Figure 18: Settings of LAS Dataset to Raster.

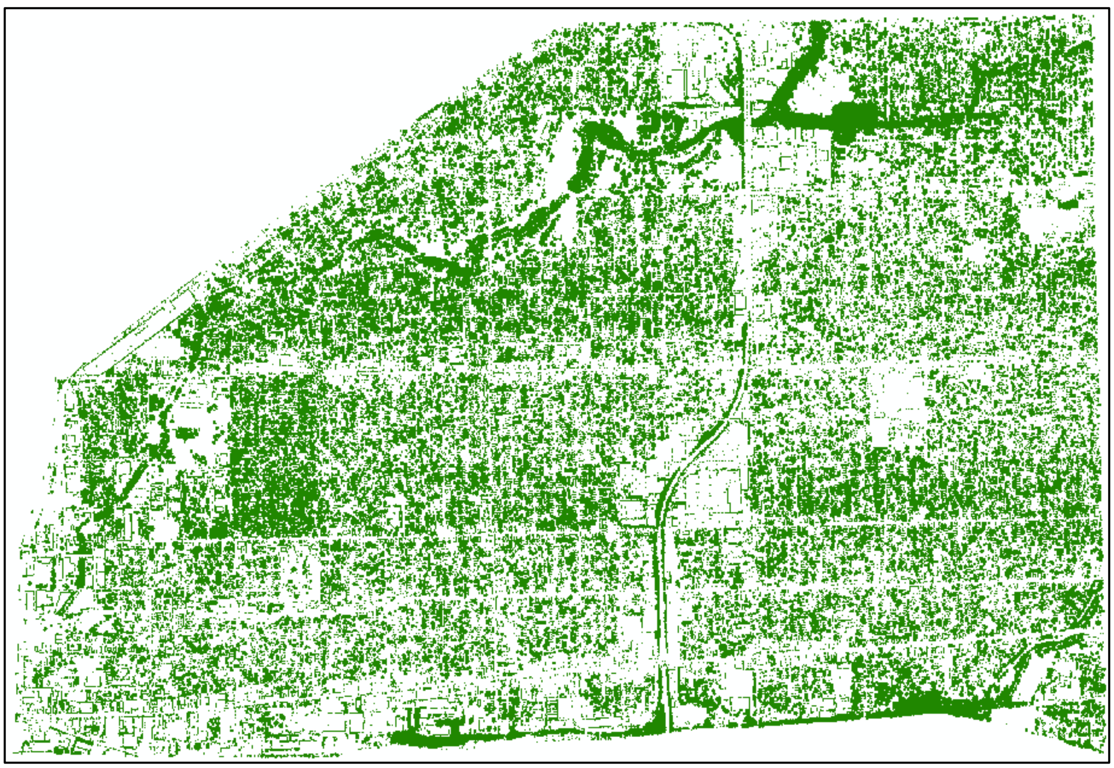


Figure 19: Tree Canopy Raster Converted from LiDAR Data.

Figure 20: Tree Canopy Raster Converted from LiDAR Data in Detail.

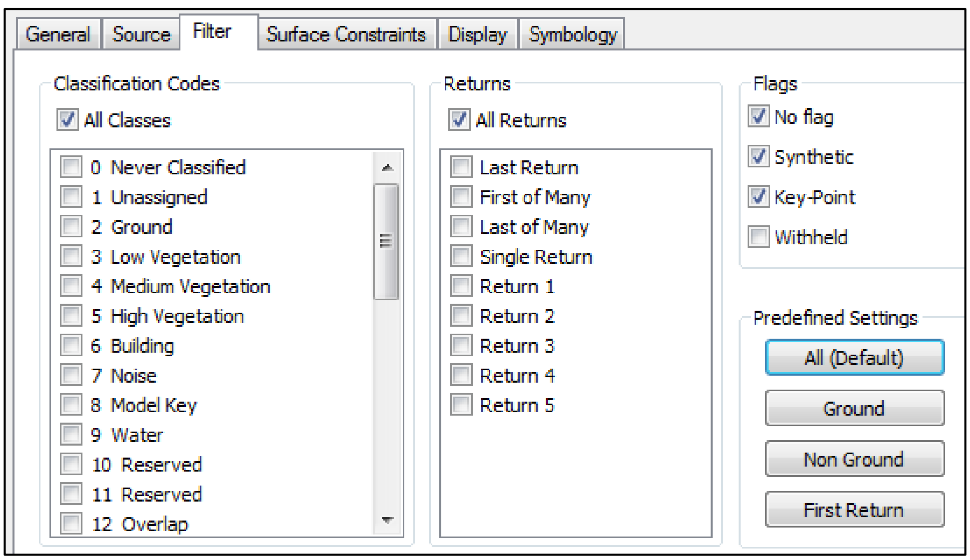


Figure 21: Setting Box of Showing DEM.

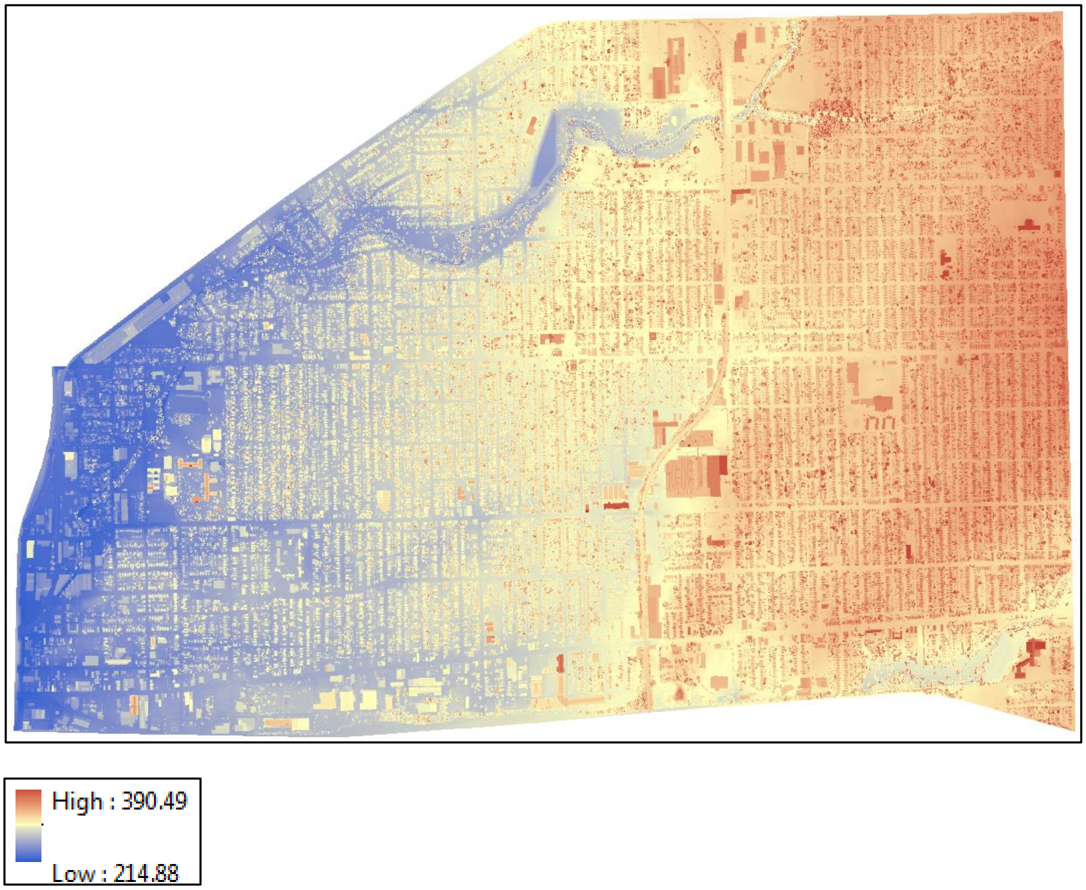


Figure 22: Layer of DEM.

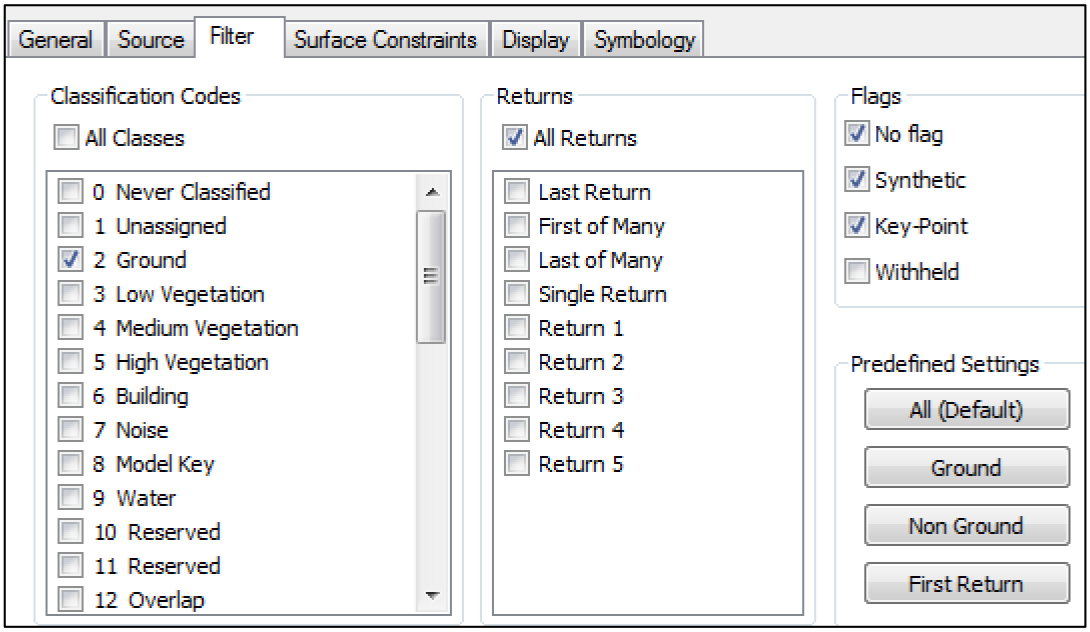


Figure 23: Setting Box of Showing Bare Ground Level (DTM).

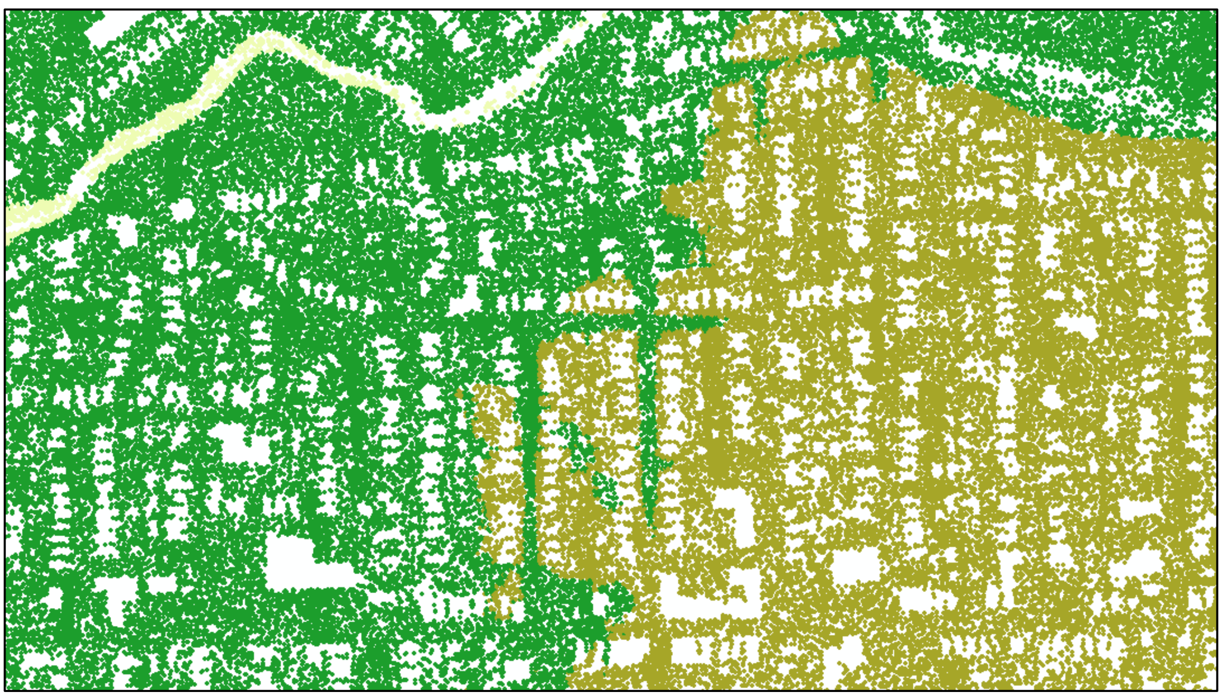


Figure 24: LiDAR Data of Bare Ground Level (DTM).

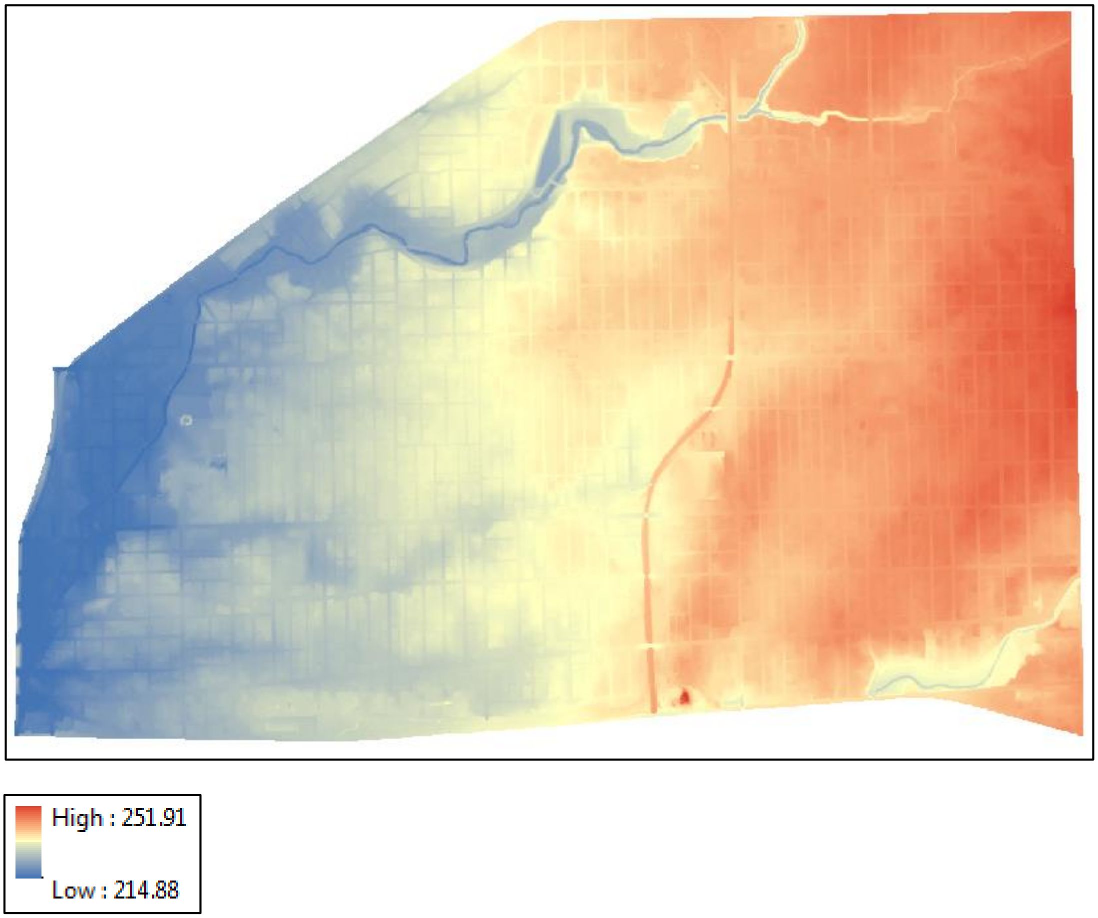


Figure 25: Bare Ground Level (DTM) Raster.

By using the Raster Calculator in ArcMap, we can get the result layer of features above the bare ground.

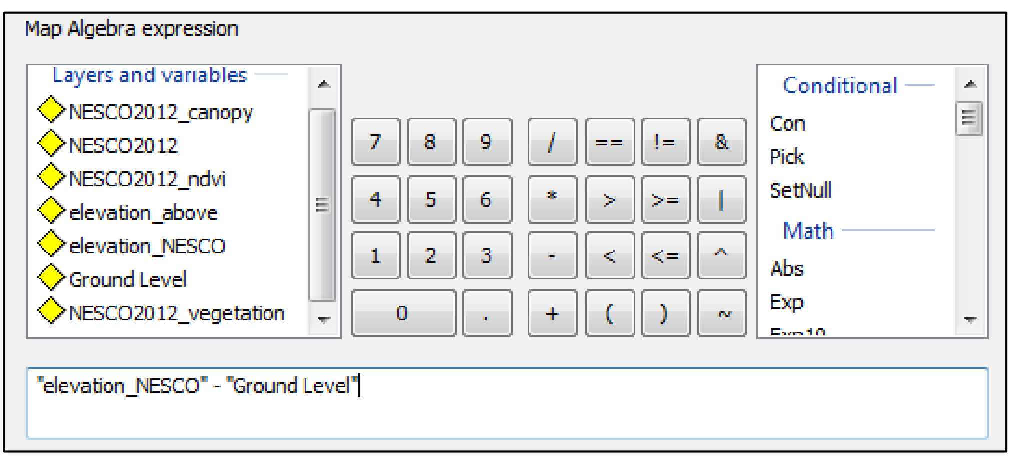


Figure 26: DEM – DTM = Features above Bare Ground Level.

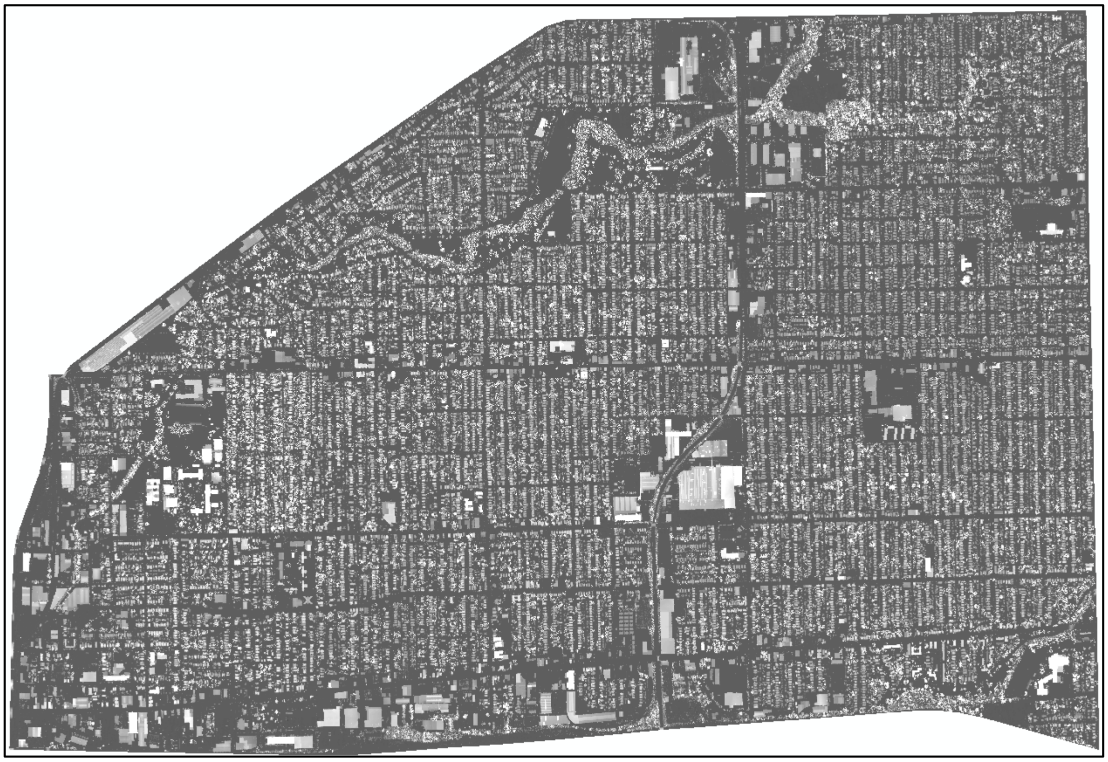


Figure 27: Layer of All Features above Bare Ground Level.

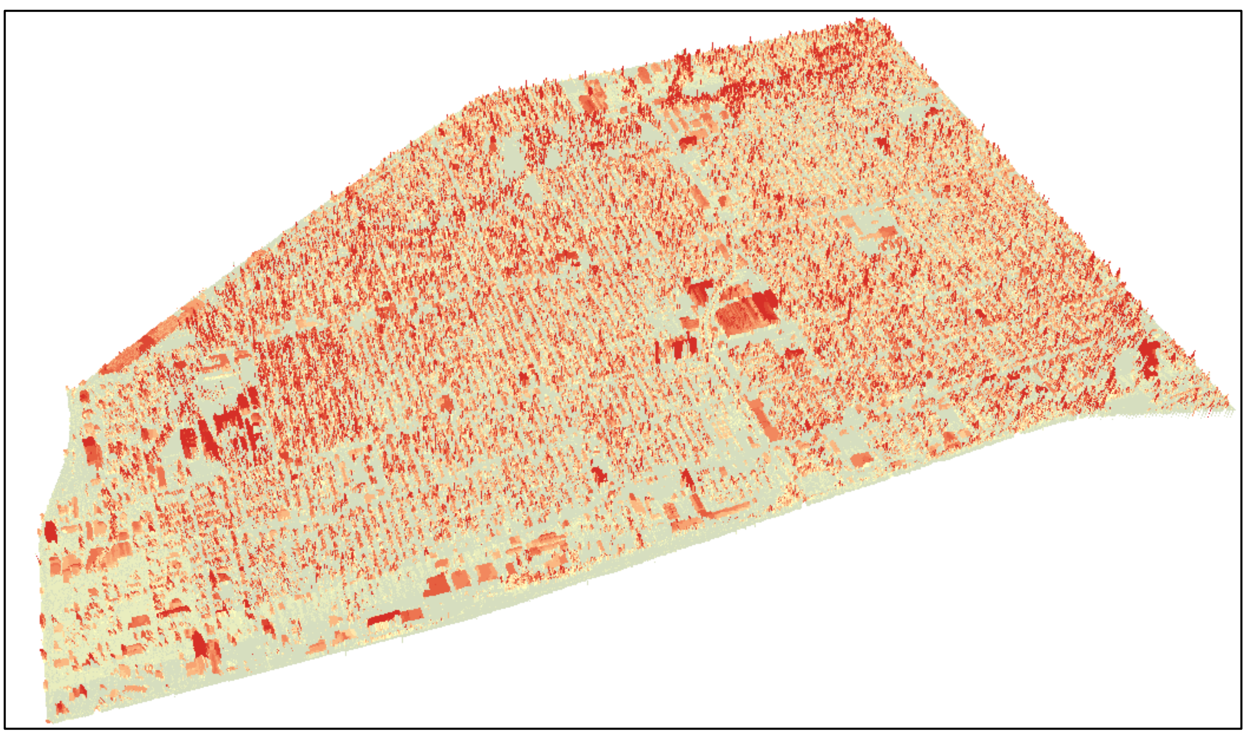


Figure 28: Layer of All Features above Bare Ground Level in 3D.

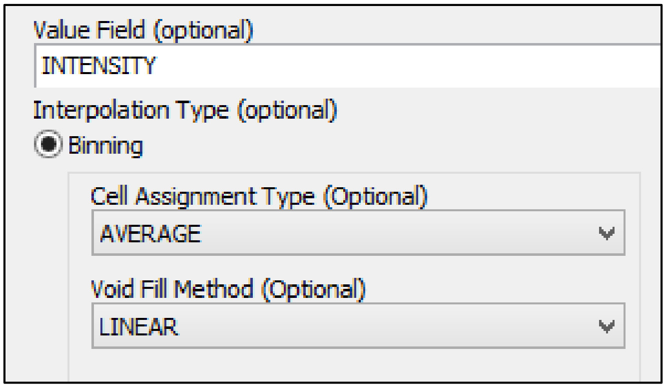


Figure 29: Settings of LAS Dataset to Raster.

Figure 30: Intensity Raster Converted from LiDAR.

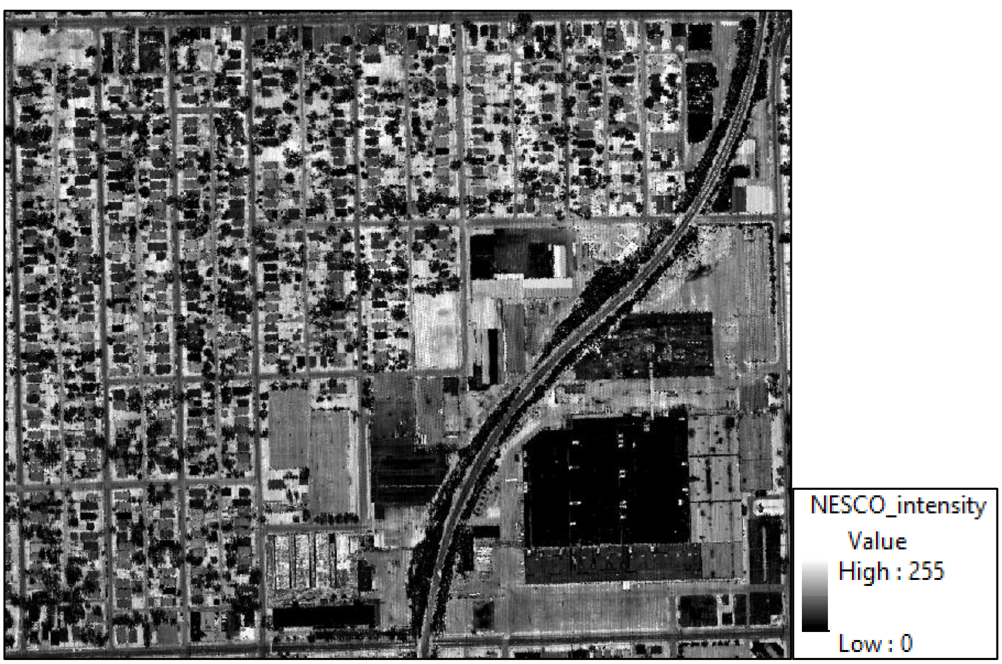


Figure 31: Intensity Raster in Detail.

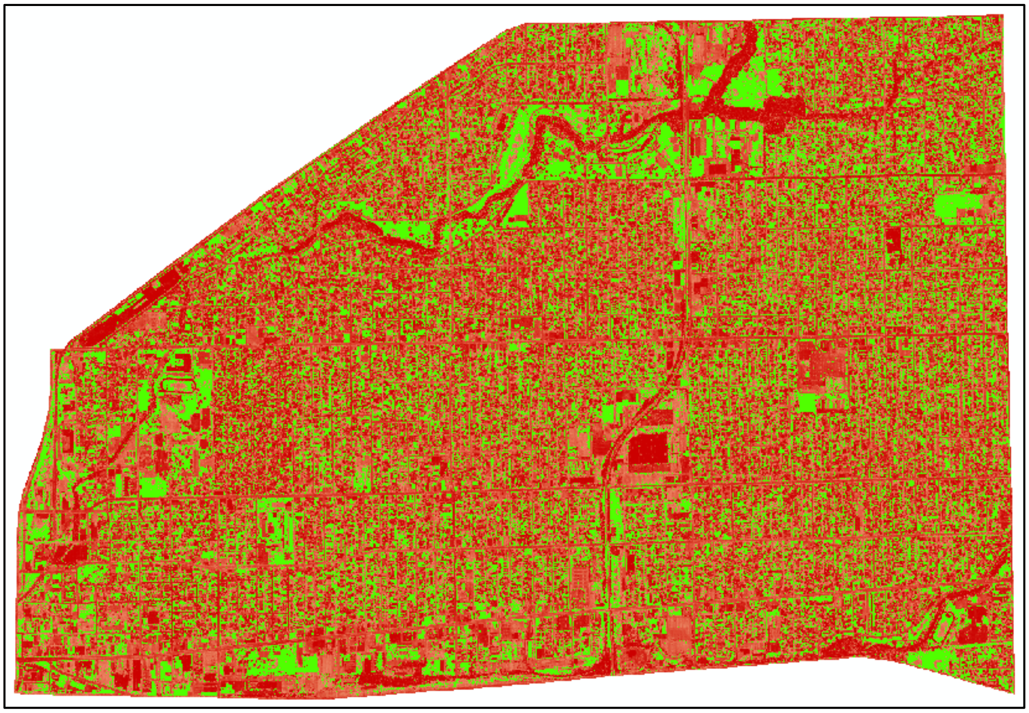


Figure 32: Intensity Raster with Different Display.

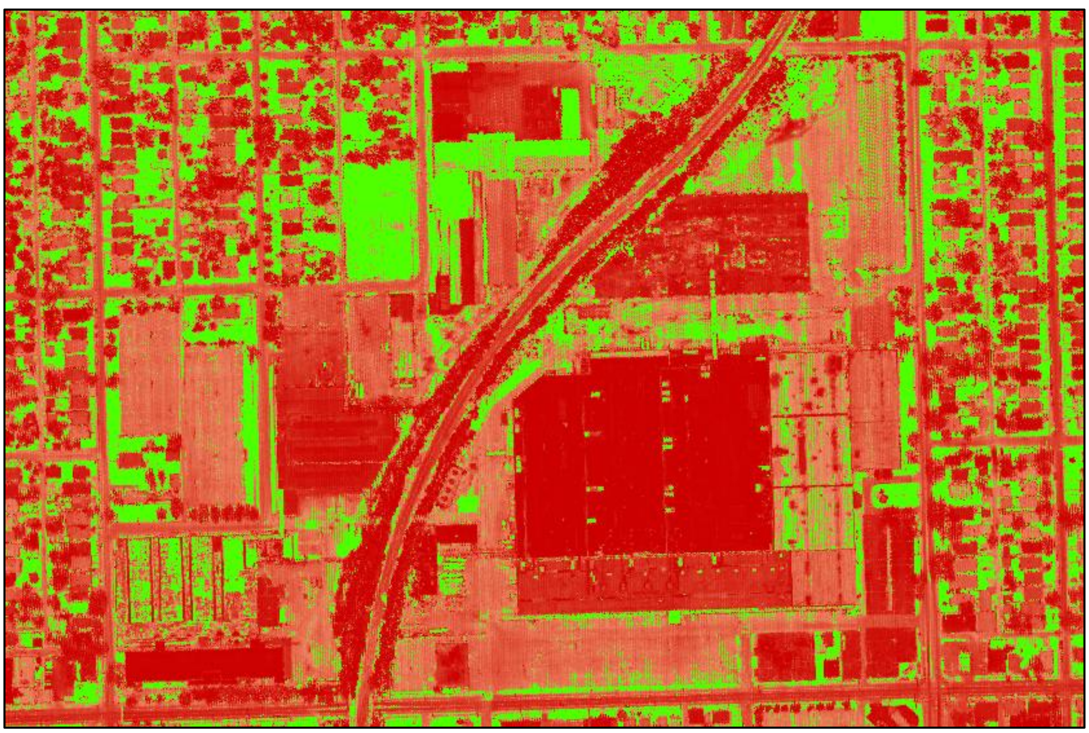


Figure 33: Intensity Raster with Different Display in Detail.

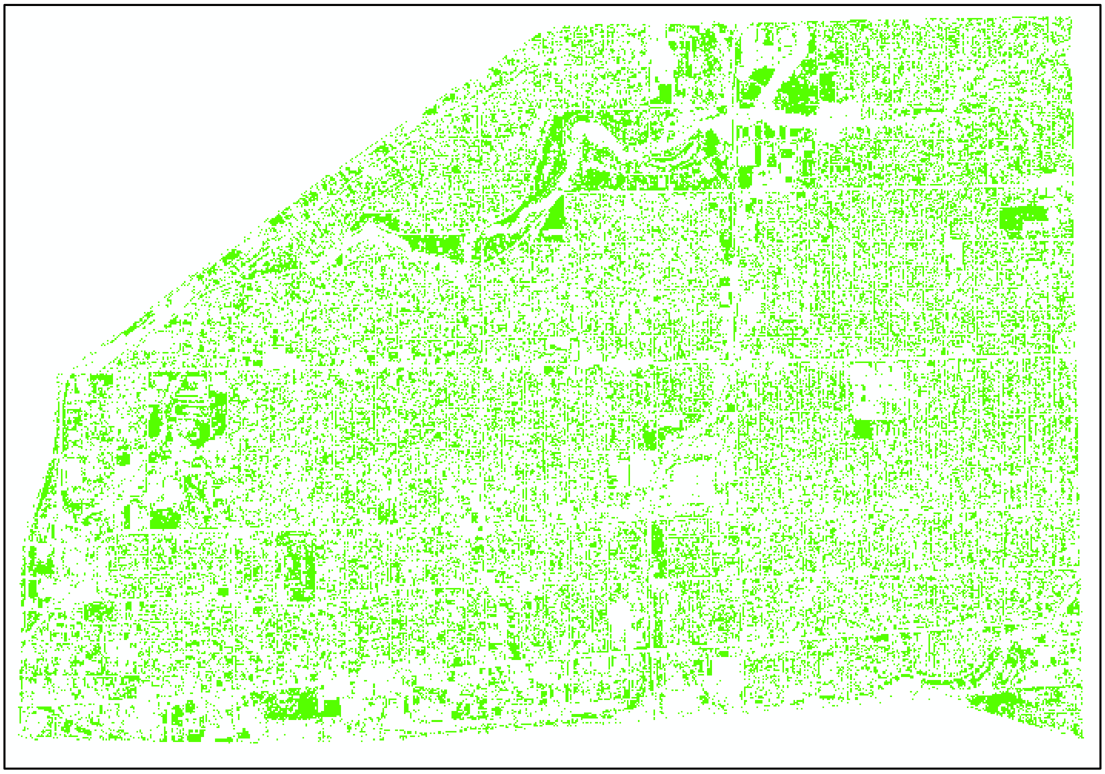


Figure 34: Intensity Raster with High Values (135 – 255).

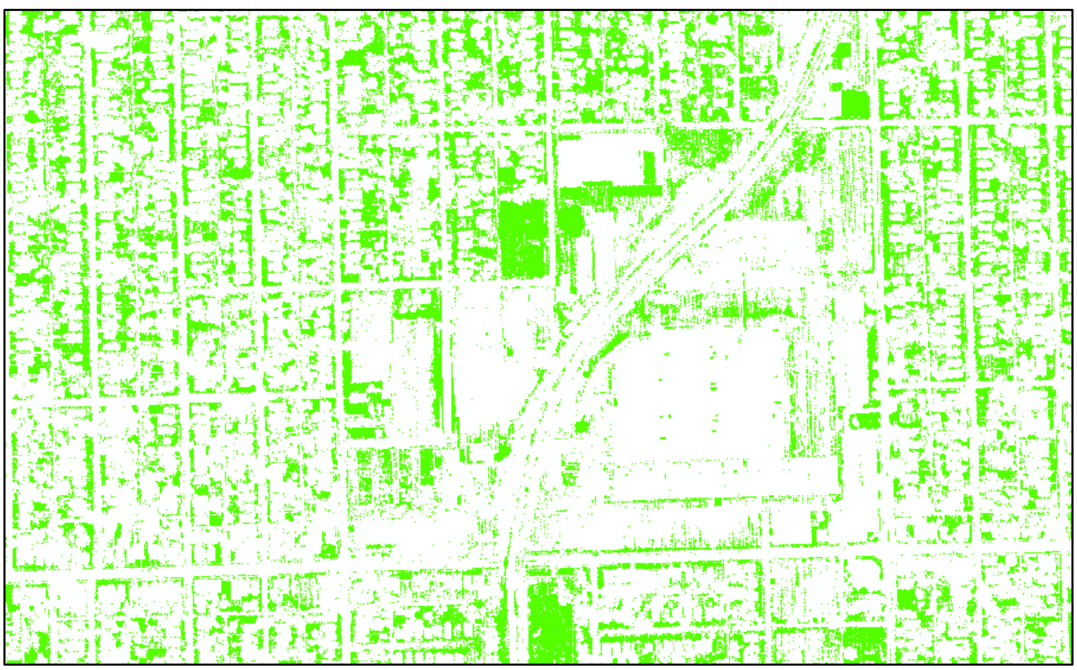


Figure 35: Intensity Raster with High Values (135 – 255) in Detail.

Figure 36: Pavements Layer including Parking Lots (Shapefile provide by IMAGIS).

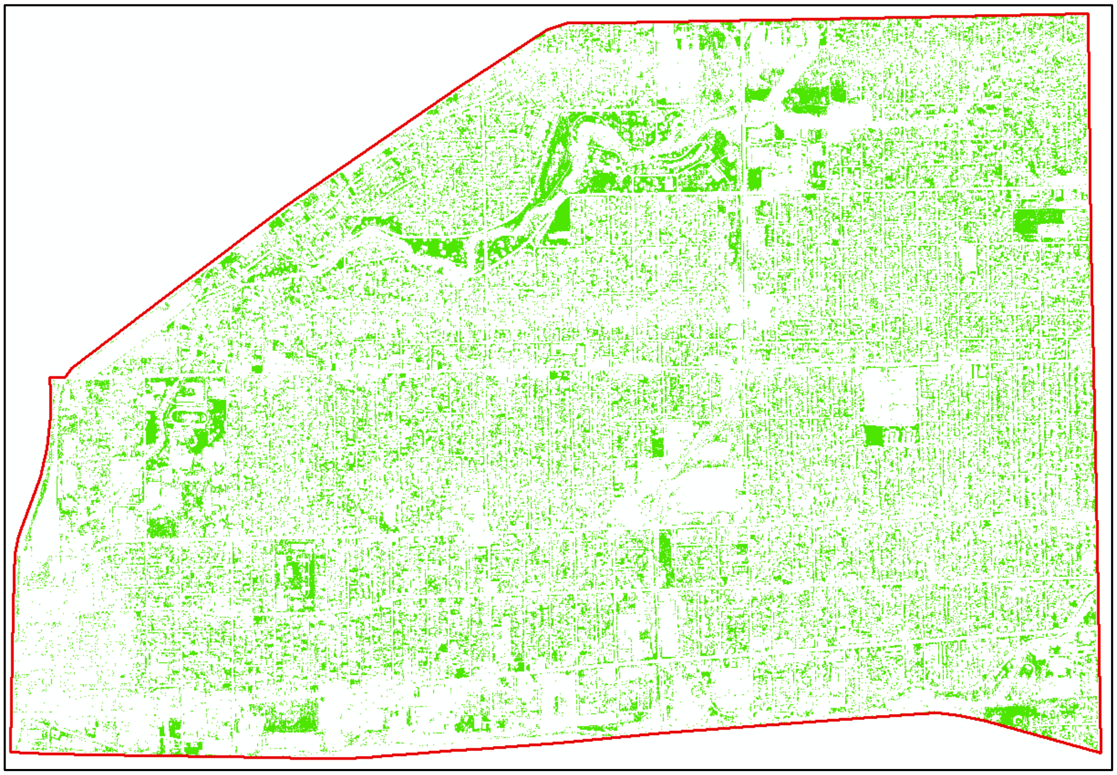


Figure 37: LiDAR Grass Classification.

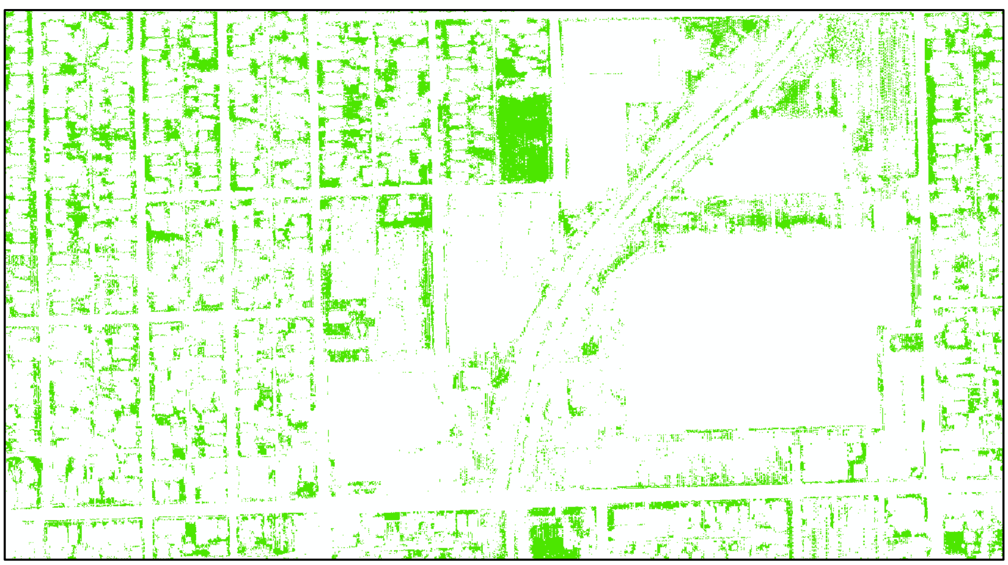


Figure 38: LiDAR Grass Classification in Detail.

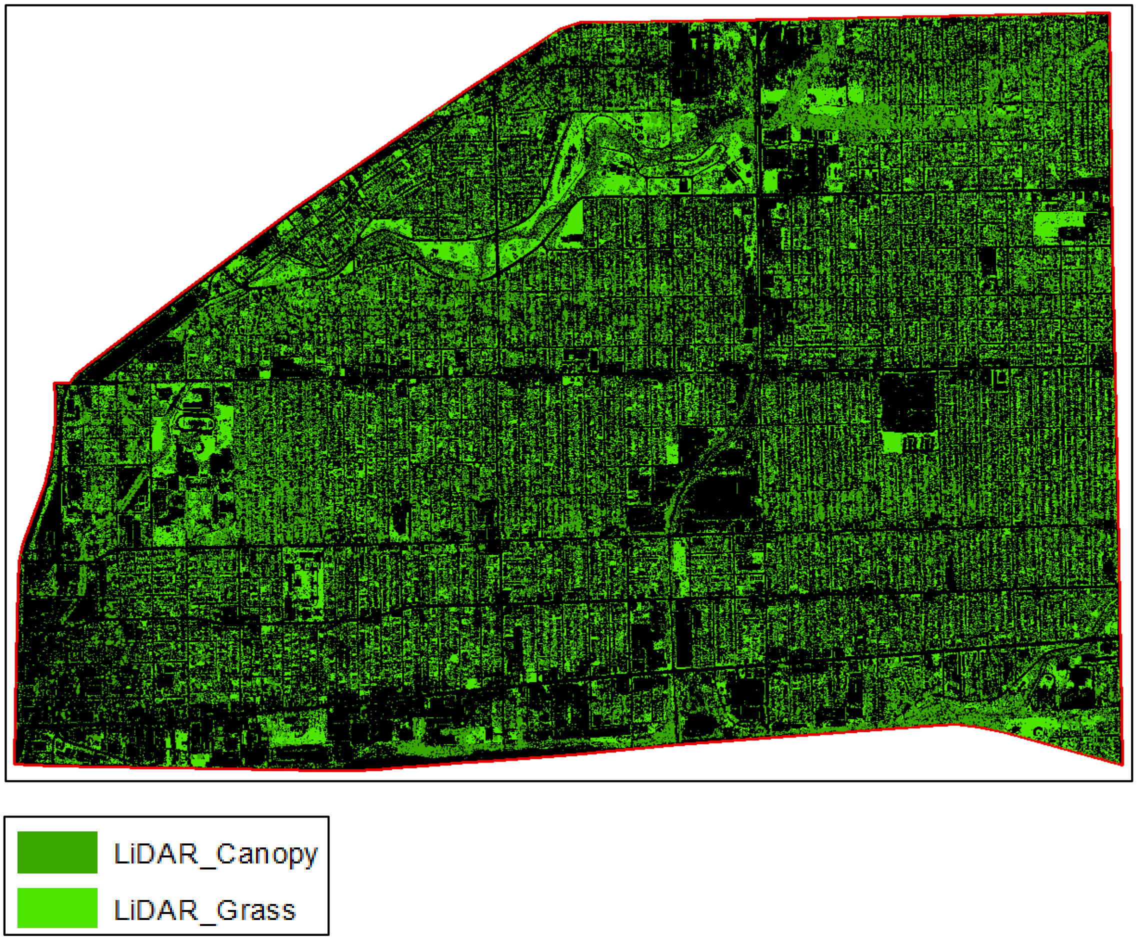


Figure 39: Result Layer of LiDAR Classification.

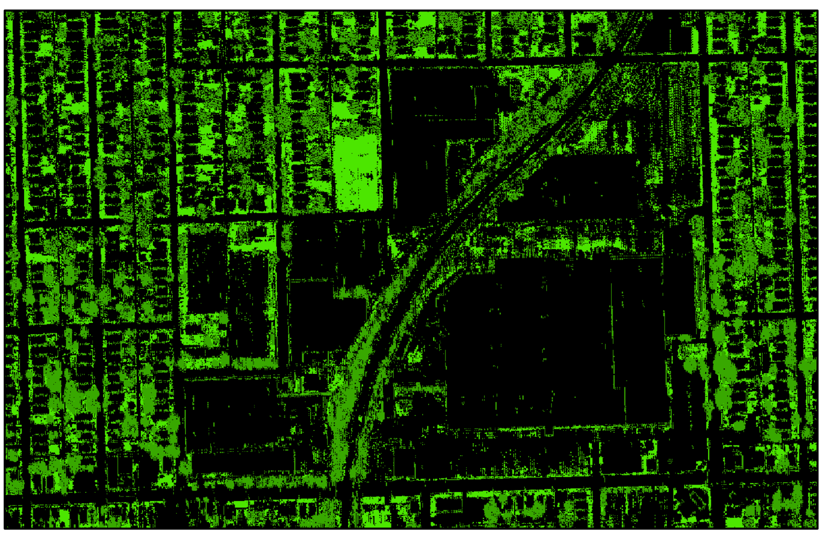


Figure 40: Result Layer of LiDAR Classification in Detail.

ACCURACY ASSESSMENT

This section will display the statistics of the results of supervised, unsupervised 100 – Class, and LiDAR classifications and the analysis of accuracy assessment. Figure 41, 42, and 43 show the result layers of these three classifications.

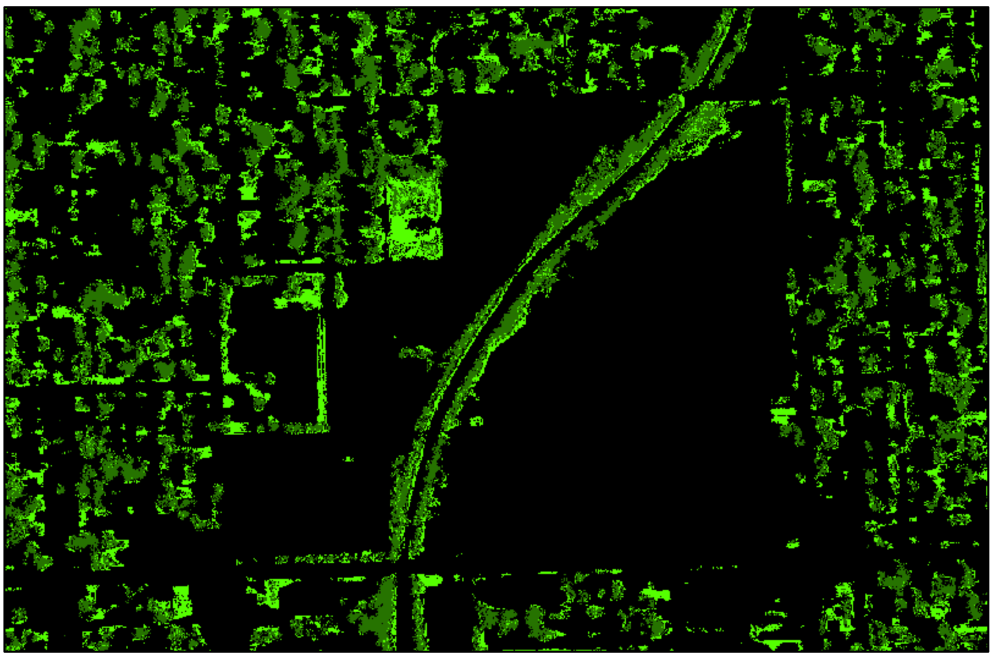


Figure 41: 100 – Class Classification of NESCO.

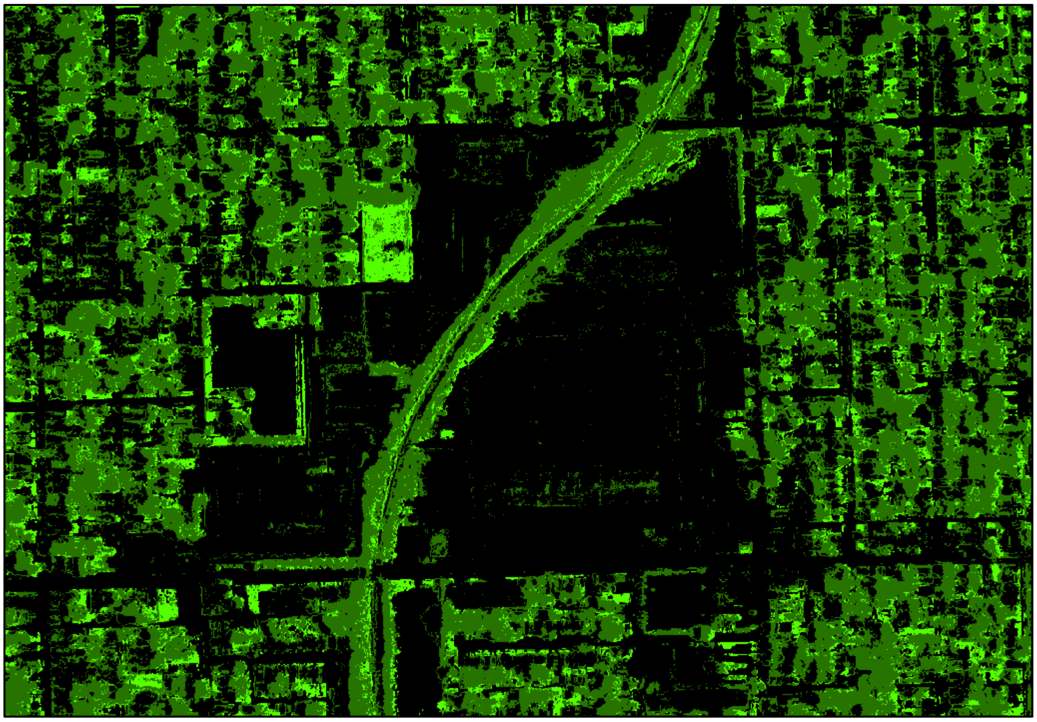


Figure 42: Supervised Classification of NESCO.

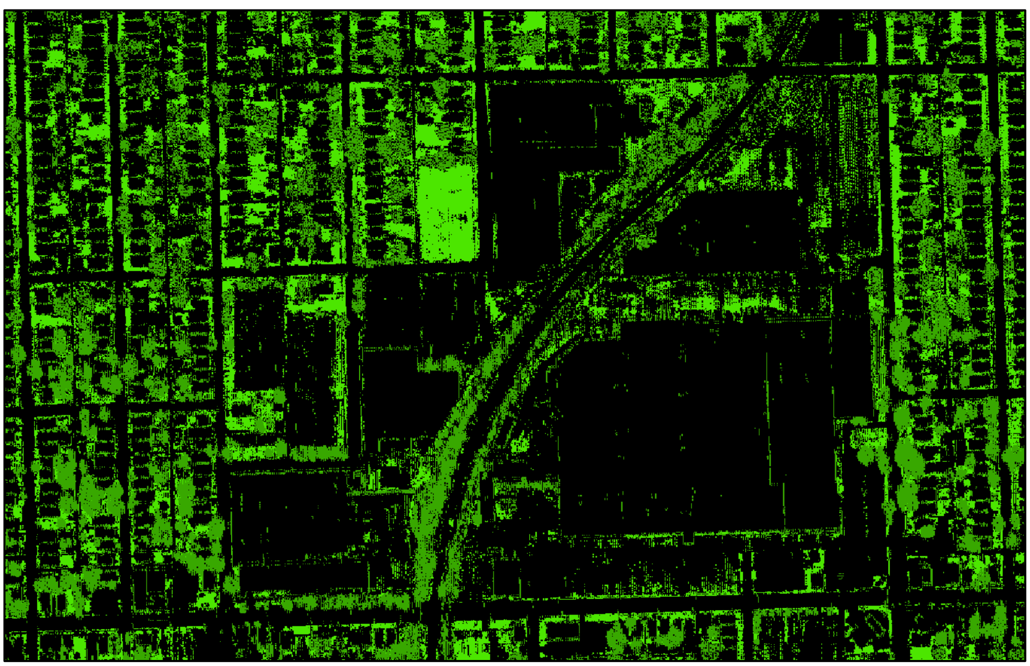


Figure 43: LiDAR Classification of NESCO.

The tables (Table 5 and 6) below show the area of classified canopy and grass areas in each classification of NESCO.

	Canopy	Grass	Unclassified	Total Area
Unsupervised	2309447	1564976	11015577	14890000
Supervised	5994194	1536745	7359061	14890000
LiDAR	2523645	3377095	8989260	14890000

Table 5: Table for Statistics of Three Types of Classification (Unit: Meter2).

	Canopy	Grass	Unclassified	Total Area
Unsupervised	0.892	0.604	4.253	5.749
Supervised	2.314	0.593	2.842	5.749
LiDAR	0.974	1.304	3.471	5.749

Table 6: Table for Statistics of Three Types of Classification (Unit: Mile2).

This research will use an Indianapolis Center Township MrSID image with 3 – band, 20 – compression, based on 6 – inch as a reference map for the accuracy assessment. This aerial photo was taken in October 2011 and provided by IMAGIS. The MrSID image is shown below (Figure 44).

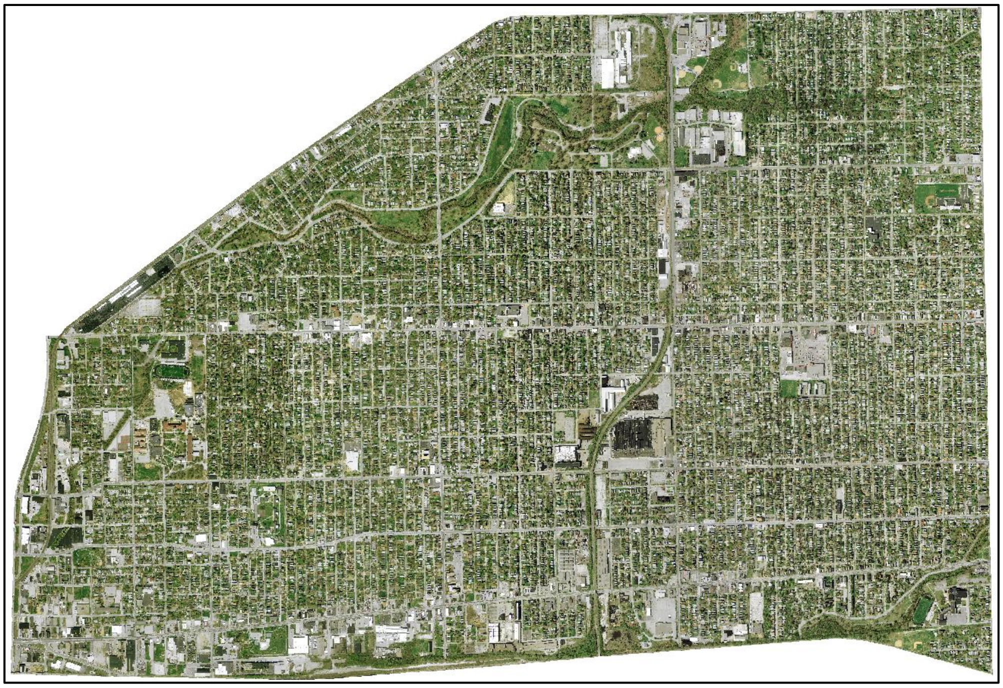


Figure 44: Indianapolis Center Township MrSID image (October 2011 from IMAGIS).

Congalton and Green (2008) stated that accuracy assessment requires an adequate number of samples per class to be collected so that the assessment would be statistically valid for representing the accuracy of the map. A general guideline or good “rule of thumb” suggests collecting a minimum of 25 samples for each class for maps of less than one million acres in size and fewer than 12 classes. Larger area maps or more complex maps should have 75 to 100 samples for each class. This guideline was empirically derived from many projects and was proved to be a good balance between statistical validity and practicality.

As to improve the quality of the accuracy assessment in this research, there will be 128 samples randomly chosen for each canopy class and grass class.

Carletta (1996) stated that the kappa coefficient measures pairwise agreement for qualitative items. $K = (P(O) – P(E)) / (1 – P(E))$. $P(O)$ is overall accuracy. $P(E)$ is expected accuracy.

There is a table of kappa interpretation (Viera & Garrett, 2005).

Interpretation of Kappa	
Kappa	Accuracy
< 0	Not accurate
0.01 – 0.20	Low accuracy
0.21 – 0.40	Fair accuracy
0.41 – 0.60	Moderate accuracy
0.61 – 0.80	Substantial accuracy
0.81 – 0.99	Almost perfect accuracy

Table 7: Table for Table of Kappa Interpretation.

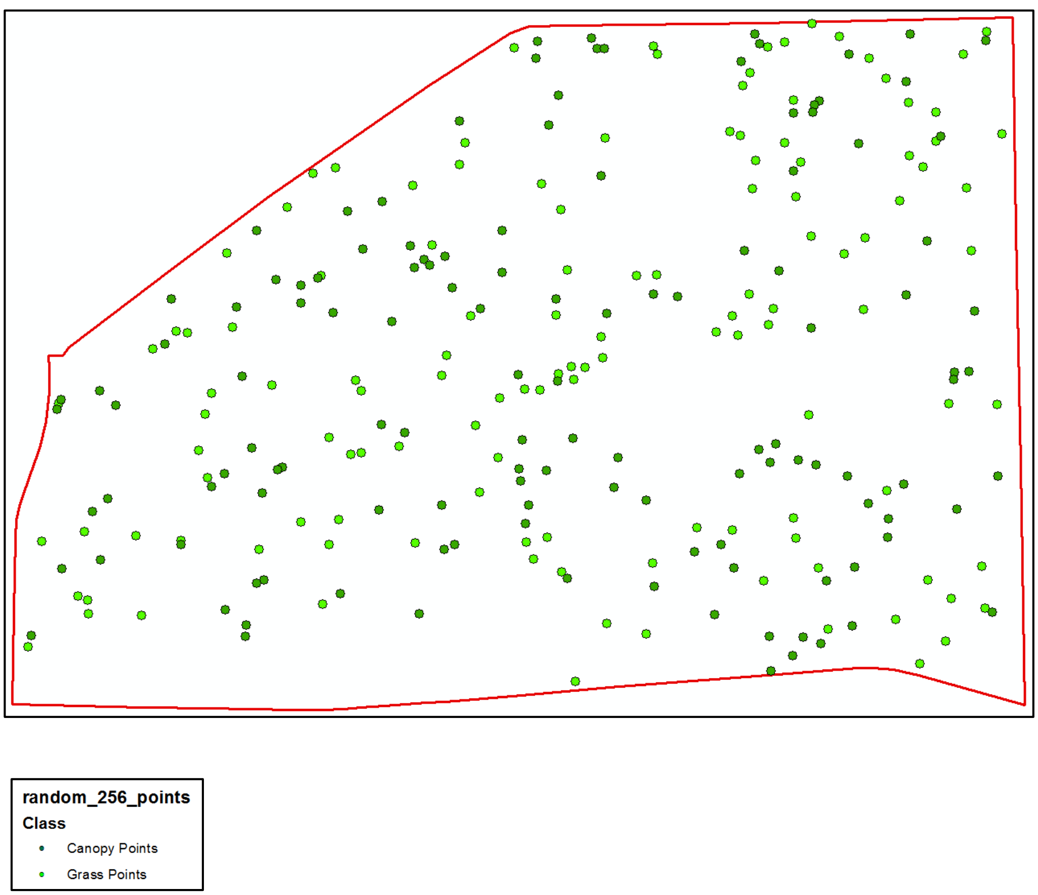


Figure 45: 256 Samples randomly chosen in Canopy Class and Grass Class.

The next step is to determine how many sample points are classified correctly in these three types of classifications. For instance, if a canopy sample point is shown in the canopy class of LiDAR classification, then it means that this canopy sample point is correctly classified by the LiDAR classification. Otherwise, it is incorrectly classified. These three tables (Table 8, 9 and 10) below are the reports of accuracy assessment for unsupervised classification, supervised classification, and LiDAR classification.

In the accuracy tables, there are two columns called producer's accuracy and user's accuracy. The producer's accuracy is the reference total divided by the number correct. The user's accuracy is the classified total divided by the number correct.

Congalton (1991) stated that producer's accuracy indicates how well a certain area can be classified while user's accuracy indicates the probability that a classified pixel on the image represents that category on the ground.

Unsupervised	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Unclassified	0	116	0		0
Canopy	128	69	54	0.421875	0.782608696
Grass	128	71	59	0.4609375	0.830985915
Totals	256	256	113		

Overall Classification (Producer's) Accuracy: 0.44140625

Kappa Statistics: 0.231183

Table 8: Accuracy Table for Unsupervised (100 – Class) Classification.

Supervised	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Unclassified	0	42	0		
Canopy	128	141	102	0.796875	0.723404255
Grass	128	73	60	0.46875	0.821917808
Totals	256	256	162		

Overall Classification (Producer's) Accuracy: 0.6328125

Kappa Statistics: 0.369127517

Table 9: Accuracy Table for Supervised Classification.

LiDAR	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy
Unclassified	0	15	0		
Canopy	128	122	122	0.953125	1
Grass	128	119	119	0.929688	1
Totals	256	256	241		

Overall Classification (Producer's) Accuracy: 0.941406

Kappa Statistics: 0.889299

Table 10: Accuracy Table for LiDAR Classification.

DISCUSSION

Classifying urban forestry is an important way to develop strategies in order to improve the environment of cities and urban areas. GIS technology provides many ways to classify urban forestry such as unsupervised, supervised, and LiDAR classification.

Unsupervised and supervised classification using satellite images and aerial photos can retrieve urban forestry information. However, the urban areas vary considerably even within one or a few square meters, plus there are many shadow areas in the satellite and aerial images, which result in many errors of unsupervised and supervised classification.

LiDAR data has horizontal and vertical information at high spatial resolutions and vertical accuracies (Lim, Treitz, Wulder, St-Onge, & Flood, 2003). The multiple – return LiDAR data can provide vegetation height information to within ±20 cm of its true height (Hodgson, Jensen, Tullis, Riordan, & Archer, 2003).

Overall, the LiDAR classification is the most accurate classification method out of these three methods discussed in this research project: unsupervised (100 – class) classification, supervised classification, and LiDAR classification. The LiDAR classification has the overall accuracy with 92.97% while the unsupervised and supervised classifications have only 44.14% and 63.28% as their overall accuracy results.

There are several advantages of the LiDAR classification model. LiDAR data has the information of elevation, the number of (Transmitted pulse) returns, and the intensity for each point, which can easily help generate the layers of each canopy and grass classifications. LiDAR classification will not be distorted by shadows, because the LiDAR image is generated with pulse points instead of being taken by a camera. There are several steps in the LiDAR classification model of this research project and each step is fairly easy to set, but the computer processing of some steps may take a while to complete since there are many times of huge calculation converting raster layers to polygon layers. The LiDAR classification model shows the most accurate result for classifying canopy and grass classes in this research project. Technically, the classification results of the LiDAR model processed by different persons should be very similar, because most of the results of each step is calculated by the computer. However, the flaw of the LiDAR grass classification is that there are a few errors on the building roofs, pavements, parking lots, driveways, and some other impervious surfaces, and those errors will decrease the classification accuracy. The future improvement of this model would be to use an accurate layer of impervious surfaces to remove those errors on impervious surfaces.

The unsupervised (100 – class) classification model is very easy and quick to perform, but this classification model is highly and negatively affected by the shadows in the aerial image and also it requires a very high resolution image to run the process. Different persons running this model may have very different classification results, because classifying the 100 classes into canopy, grass, and unsupervised classes is based on the persons' choices.

Likewise, the supervised classification model is also very easy and quick to perform. Again, it is also highly and negatively affected by the shadows in the aerial image and it requires a very high resolution image to get an accurate classification. Like the unsupervised classification model, the supervised classification model may generate different results as different people run this model, because different people would choose different "training areas" that would result in different classification results.

CONCLUSION

The total area of NESCO is 5.749 Miles^2. The canopy, grass, and unclassified areas of unsupervised (100 – class) classification are 0.892 Miles^2, 0.604 Mile^2, and 4.053 Mile^2; areas of supervised classification are 2.314 Mile^2, 0.593 Mile^2, and 2.842 Mile^2; areas of LiDAR classification are 0.974 Mile^2, 1.304 Mile^2, and 3.471 Mile^2.

In the accuracy assessment discussed above, there were total of 256 randomly chosen points tested. From the accuracy tables, the results of the canopy, grass, and overall accuracy of unsupervised (100 – class) classification are 42.19%, 46.09%, and 44.14%; the results of supervised classification are 79.69%, 46.88%, and 63.28%; the results of LiDAR classification are 95.31%, 92.97%, and 94.14%.

Clearly, the LiDAR classification is much more accurate than the other two classification methods in this research project.

REFERENCES

Ahmad, A., & Quegan, S. (2012, November). Analysis of Maximum Likelihood classification technique on Landsat 5 TM satellite data of tropical land covers. In *Control System, Computing and Engineering (ICCSCE), 2012 IEEE International Conference on* (pp. 280 – 285). IEEE.

Al-Ahmadi, F. S., & Hames, A. S. (2009). Comparison of four classification methods to extract land use and land cover from raw satellite images for some remote arid areas, kingdom of Saudi Arabia. *Earth*, *20*(1), 167 – 191.

Arai, K., & Bu, X. Q. (2007). ISODATA clustering with parameter (threshold for merge and split) estimation based on GA: Genetic Algorithm. *Reports of the faculty of science and engineering, Saga University, 36*(1), 17 – 23.

Ari, C., & Aksoy, S. (2010, July). Unsupervised classification of remotely sensed images using gaussian mixture models and particle swarm optimization. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International* (pp. 1859 – 1862). IEEE.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, *22*(2), 249 – 254.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of Environment*, *37*(1), 35 – 46.

Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.

Coder, K. D. (1996). Identified benefits of community trees and forests. *University of Georgia*.

Dwyer, J. F., McPherson, E. G., Schroeder, H. W., & Rowntree, R. A. (1992). Assessing the benefits and costs of the urban forest. *Journal of Arboriculture*, *18*, 227 – 227.

Imagine, E. R. D. A. S. (1999). Tour Guides V8. 4, ERDAS. *Inc., Atlanta Georgia*.

Gatziolis, D., & Andersen, H. E. (2008). *A guide to LiDAR data acquisition and processing for the forests of the Pacific Northwest* (pp. 1 – 32). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Harrap, R., & Lato, M. (2006). *An Overview of LIDAR for Urban Applications*. Technical report. *Queen's University.*

Hodgson, M. E., Jensen, J. R., Tullis, J. A., Riordan, K. D., & Archer, C. M. (2003). Synergistic use of lidar and color aerial photography for mapping urban parcel imperviousness. *Photogrammetric Engineering and Remote Sensing*, *69*(9), 973 – 980.

Isenburg, M. (2013). LASzip: lossless compression of LiDAR data. *Photogrammetric Engineering and Remote Sensing*, *79*(2), 209 – 217.

Kokalj, Ţ., & Oštir, K. (2007). Land cover mapping using Landsat satellite image classification in the Classical Karst – Kras region, *36*(3), 433 – 440.

Liu, X. (2008). Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography*, *32*(1), 31 – 49.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in physical geography*, *27*(1), 88 – 106.

Palmer, T. C., & Shan, J. A. (2002). Comparative study on urban visualization using LIDAR data in GIS. *URISA Journal*, *14*(2), 19 – 25.

Renslow, M., Greenfield, P., & Guay, T. (2000). Evaluation of multi – return LIDAR for forestry applications. *US Department of Agriculture Forest Service – Engineering, Remote Sensing Applications. http://www. ndep. gov/USDAFS/LIDAR.pdf*

Rowntree, R. A., & Nowak, D. J. (1991). Quantifying the role of urban forests in removing atmospheric carbon dioxide. *Journal of Arboriculture*, *17*(10), 269 – 275.

Shan, J., & Sampath, A. (2005). Urban DEM generation from raw LIDAR data: a labeling algorithm and its performance. *Photogrammetric Engineering & Remote Sensing*, *71*(2), 217 – 226.

Singh, K. K., Vogler, J. B., Meng, Q., & Meentemeyer, R. K. (2010, June). Mapping land use patterns in an urbanizing landscape using LiDAR intensity data. In *Optical Remote Sensing of the Environment*. Optical Society of America.

Veneziano, D., Hallmark, S., & Souleyrette, R. (2002). Accuracy Evaluation of LIDAR – Derived Terrain Data for Highway Location. *Computer – Aided Civil and Infrastructure Engineering*.

Verma, R., Kumari, K. S., & Tiwary, R. K. (2009). Application of Remote Sensing and GIS Technique for Efficient Urban Planning in India. In *Geomatrix Conference Proceedings*.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, *37*(5), 360 – 363.

Yang, L., Xian, G., Klaver, J. M., & Deal, B. (2003). Urban land – cover change detection through sub – pixel imperviousness mapping using remotely sensed data. *Photogrammetric engineering and remote sensing*, *69*(9), 1003 – 1010.

CURRICULUM VITAE

Nan Ye

Education

M.S. (January 2014), Geographic Information Science (GIS), Indiana University, Indianapolis, IN

B.S. (May 2010), Survey and Mapping Engineering, East China Institute of Technology, Fuzhou, Jiangxi

Coursework:

- GIS
- Advanced GIS
- Remote Sensing
- Advanced Remote Sensing
- Computer Cartography& Graphics
- Applied Spatial Statistics
- Transportation Analysis
- GIS Seminar
- Contemporary Health Issues (Fall 2013)

Professional Experience

- Research at Geography Department IUPUI working on Near East Side Community Organization Plantable Area project for Keep Indianapolis Beautiful, Inc. (September 2012 – December 2013)

- Research at Geography Department IUPUI collaborating on Extreme Heat Vulnerability Maps for Chicago, Dayton, Indianapolis, Philadelphia and Phoenix. (May 2013 – December 2013)
- Internship at The Polis Center IUPUI updating critical infrastructure GIS databases for the Indiana Department of Homeland Security. (May 2012 – August 2012)

Skills

- Advanced knowledge of using ArcGIS, ArcGIS Server, ERDAS Imagine, JavaScript API for ArcGIS, and TransCAD
- Advanced knowledge of python scripting for ArcGIS.
- Advanced knowledge of spatial analysis, transportation networks
- Experience creating databases and queries in Microsoft Access
- Advanced knowledge of Microsoft Office, Digital Image Pro